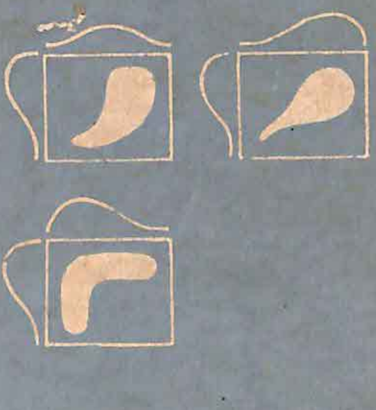
শি

পরি

দুর্গা

ক্ষায়
সংখ্যান



দাস ভট্টাচার্য্য

শিক্ষায় পরিসংখ্যান

অধ্যাপক দুর্গাদাস ভট্টাচার্য
এম্, এস্-সি ; বি, টি.

জ্ঞানেন্দ্রমোহন সেন রিসার্চ স্কলার কলিকাতা বিশ্ববিদ্যালয়,
শিক্ষক শিক্ষণ বিভাগ কল্যাণী বিশ্ববিদ্যালয়, ভূতপূর্ব
রিসার্চ ফেলো, পরিসংখ্যান বিভাগ, কলিকাতা
বিশ্ববিদ্যালয় এবং দার্জিলিং শ্রীরামকৃষ্ণ
বি, টি, কলেজের ভূতপূর্ব অধ্যাপক।

দি
নিউ বুক ষ্টল
পুস্তক বিক্রেতা ও প্রকাশক
৫।১ রমানাথ মজুমদার স্ট্রীট,
কলিকাতা-৯

প্রকাশক :
শ্রীমহেন্দ্রনাথ পাল
৬, রমানাথ মজুমদার স্ট্রীট
কলিকাতা-৯

মূল্য—৮·০০

মুদ্রাকর :
শ্রীরামচন্দ্র দে
ইউনাইটেড আর্ট প্রেস
২৫বি, হিদারাম ব্যানার্জী লেন
কলিকাতা-১২

সূচীপত্র

A. Introduction :

শিক্ষায় পরিসংখ্যান

উপক্রমণিকা

'পরিসংখ্যান' শব্দটি দুই প্রকারে ব্যবহৃত হয়। প্রথমটি হইতেছে ঘটনাসমূহকে সংখ্যা দ্বারা প্রকাশ এবং দ্বিতীয়টি হইতেছে ঐ সকল সংখ্যাকে ঠিক মত ব্যবহারোপযোগী করার নীতি বা পদ্ধতি। ঘটনানির্দেশক সংখ্যা সংগ্রহ, উপস্থাপন, বিশ্লেষণ ও তাহার ব্যাখ্যাই হইতেছে পরিসংখ্যান। তাই ঘটনাসমূহকে কেবলমাত্র সংখ্যাদ্বারা প্রকাশই যথেষ্ট নহে; সেইগুলিকে পর্যায়ক্রমে সজ্জিত করা, সুসংবদ্ধ করা, তাহাকে বিভিন্নভাবে তুলনা করা, প্রাপ্ত প্রতি তথ্যকে বিশ্লেষণ করা একান্ত প্রয়োজন। শুধু তাহাই নহে। বিশ্লেষণের পরে তাহার ব্যাখ্যা ও সমালোচনা করার ক্ষমতার উপরেই প্রথমোক্ত কার্যগুলির নির্ভরযোগ্যতা নির্ভর করিতেছে। তাই পরিসংখ্যানের গণ্ডী কেবলমাত্র সংখ্যাগণনার মধ্যে সীমাবদ্ধ করিলে ভুল হইবে। ইহার এলাকা আরও ব্যাপক ও বিস্তৃত।

পরিসংখ্যানকে পদার্থবিদ্যা, রসায়নবিদ্যা, অর্থনীতি বা সমাজ বিজ্ঞানের মত কোন বিষয় হিসাবে গণ্য করা উচিত নয়। পরিসংখ্যান বিজ্ঞান নহে; উহা একটি বৈজ্ঞানিক পদ্ধতি। পরিসংখ্যানে ন্যূনতম জ্ঞান না থাকিলে সমাজ বিজ্ঞান অধ্যয়ন অন্ধ মানুষের বিড়ালশূন্য অন্ধকার ঘরে কালো বিড়াল খুঁজিবার চেষ্টার মতই ব্যর্থ হইবে। ক্রমবর্ধমান জটিল সমাজ ব্যবস্থার মধ্যে সমস্যাজর্জরিত মানুষের কাছে পরিসংখ্যানপদ্ধতি তাই আজ নিত্যপ্রয়োজনীয় বন্ধু হিসাবে পরিগণিত হইয়াছে।

পরিসংখ্যান বা Statistics-এর ব্যুৎপত্তিগত অর্থের মধ্যে তাহার জন্ম হইয়াছে। যুদ্ধ বা স্বাভাবিক অবস্থাতে রাজ্য সমূহের (States) পরিচালনার ব্যাপারে লোক গণনা বা সম্পত্তির হিসাব অথবা তাহাদের বিশ্লেষণ একান্ত প্রয়োজনীয় হইয়া উঠিয়াছিল। পরে সমাজ ব্যবস্থা বা রাজ্য পরিচালনার বিভিন্ন ক্ষেত্রে তাহার প্রয়োগ ধীরে ধীরে বিস্তৃতি লাভ করিতে থাকে। এখন জীবনবীমা, জীববিদ্যা, অর্থনীতি, সমাজবিজ্ঞান, ব্যবসা, বাণিজ্য, কৃষি, চিকিৎসাবিদ্যা, মনস্তত্ত্ব, শিক্ষা প্রভৃতি ক্ষেত্রে তাহার ব্যাপক প্রয়োগ পরিদৃষ্ট হয়। এমনকি আজ কলাবিদ্‌, অভিনেতা, গায়ক বা ঔপন্যাসিকও মানুষের আগ্রহ, ইচ্ছা বা বিশেষ ঝোঁকের দিকে লক্ষ্য রাখার জন্য পরিসংখ্যানের সাহায্য গ্রহণ করেন।

পরিসংখ্যানের সংগা নির্ণয়ের সময় পূর্বে বলা হইয়াছে যে কোন ঘটনাকে প্রথমে সংখ্যাগতভাবে সংগ্রহ করা হয়—পরে তাহাকে উপস্থাপিত এবং বিশ্লেষিত করিয়া ব্যাখ্যা করা হয়। এই চারটি পদ্ধতি নীচে আলোচিত হইতেছে।

(১) সংখ্যা সংগ্রহ—পরিসংখ্যানে data প্রকাশিত বা অপ্রকাশিত ঘটনা হইতে গ্রহণ করা যাইতে পারে। সরকারী বিজ্ঞপ্তি, ব্যবসা সংক্রান্ত তথ্য, গবেষণাগার, ম্যাগাজিন, সংবাদ পত্র বা ব্যক্তিগত গবেষণার ফল হইতে data সংগ্রহ করা যাইতে পারে।

(২) নিজের জন্য হউক বা পরের ব্যবহারের জন্যই হউক data সংগ্রহ করিয়া তাহাদিগকে নির্দিষ্ট আকারে উপস্থাপিত করিতে হয়। সাধারণতঃ তাহাদিগকে তালিকাভূক্ত করা হয় বা লেখচিত্রে প্রকাশ করা হয়। সাধারণ অবস্থায় data গুলি সাধারণের কাছে বোধগম্য নাও হইতে পারে। কিন্তু তালিকাভূক্ত করিলে বা চিত্রে প্রকাশ করিলে অতি সাধারণ ব্যক্তিও ঐ data সম্বন্ধে কিছুটা ধারণা করিতে পারে।

(৩) বিশ্লেষণ করার কালে data গুলিকে প্রয়োজনীয় এবং যুক্তিপূর্ণ বিভাগে শ্রেণীবদ্ধ করিতে হয়। Data গ্রহণের সময় কি কি বিভাগ সম্ভব তাহা ঠিক করিয়া data লইতে হয়। পরে ঐ বিভক্ত অবস্থাতে তাহাদের গণনা কার্য করা হয়। মনে রাখা দরকার যে চিত্রে প্রকাশ করার পূর্বে যদি ঐ বিভাগ না হয়, তবে ঐ চিত্র নিরর্থক হইয়া দাঁড়াইবে।

(৪) পরিসংখ্যানের সর্বশেষ কাজ হইতেছে যে পূর্ব সংগৃহীত, উপস্থাপিত এবং বিশ্লেষিত data-এর ব্যাখ্যা করা। বিশ্লেষিত ফল হইতে আমরা কি সিদ্ধান্তে উপনীত হইতে পারি? পূর্ব পরিকল্পিত ধারণা কি data-এর ফল দ্বারা পরীক্ষিত সত্যে পরিণত হইতেছে—না ঐ data অন্য কোন নূতন ধারণার অবতারণা করিতেছে? মনে রাখা উচিত যে পরীক্ষিত এলাকার মধ্যেই যেন সিদ্ধান্ত সীমাবদ্ধ থাকে। সে সিদ্ধান্ত যে নিখুঁত, নির্ভুল বা অপরিবর্তনীয় এ ধারণা যেন পরীক্ষকের কখনও না থাকে।

পরিসংখ্যানে ভবিষ্যৎবাণী করার স্থান বিদ্যমান এবং এই ক্ষমতার জন্যই এই পদ্ধতির এত ব্যবহার। কিন্তু তাহারও একটি সীমা আছে। এই ভবিষ্যৎ-বাণী সম্ভাবনার উপর প্রতিষ্ঠিত। ইহা একেবারে নির্ভুল বা অবশ্যম্ভাবী নয়। বৃহত্তর ক্ষেত্র যখন আমাদের সাধ্যের বাহিরে দাঁড়ায় বা যখন আর্থিক ব্যয়

আমাদের এলাকাকে সংকীর্ণ করিতে বাধ্য করায়—তখন ঐ স্বল্পপরিসর স্থানের data সংগ্রহ করিয়া তাহা হইতে গোটা ক্ষেত্রের ভবিষ্যৎবাণী করা হয়। ইহার মধ্যে ভুল তো কিছু থাকিবেই। কিন্তু এই ভুলেরও একটা সীমা আছে। তাই ঠিক মত data সংগৃহীত হইলে যে ভুল হইবে তাহা হইবে সীমিত ভুল। পরিসংখ্যানবিদের মনকে তাই সংকীর্ণ গণ্ডীর মধ্যে না রাখিয়া বিস্তৃত করিতে হয়। অভ্রান্ত সত্য নয়—সম্ভাবনাপূর্ণ তথ্যের উপর দাঁড়াইয়া বিভিন্ন বিজ্ঞানের অগ্রগতির পথ প্রদর্শকের ভূমিকা গ্রহণ করিয়াছে—এই পরিসংখ্যান পদ্ধতি।

শিক্ষা ও মনোবিজ্ঞানে সংশ্লিষ্ট আমরা, আমাদের সীমিত গণ্ডীর মধ্যে শিক্ষায় ও মনোবিজ্ঞানে পরিসংখ্যানকে কতটা প্রয়োগ করিতে পারি তাহাই আমাদের বিশেষ বিবেচ্য বিষয়। পাঠদানে আমাদের শ্রম কতটা সফল হইল—শিশু আমাদের অধ্যাপনা কতটা কাজে লাগাইতে পারিল—পরিসংখ্যান তাহা নির্ণয়ে কিছুটা সাহায্য করিতে পারে। কোন বিশেষ পদ্ধতি গতানুগতিক পদ্ধতির তুলনায় কতটা সফলতাদান করিতে পারে বা আদৌ পারে কিনা—তাহার তুলনামূলক আলোচনা পরিসংখ্যানে পাই। পরবর্তী আলোচনা সমূহকে বিশ্লেষিত করিয়া শিক্ষা ও মনস্তত্ত্বে কিভাবে পরিসংখ্যান পদ্ধতি কাজে লাগান যাইতে পারে তাহার সংক্ষিপ্ত বিবরণ নীচে প্রদত্ত হইল।

(ক) শিক্ষায় বা মনস্তত্ত্বে প্রাপ্ত সংখ্যাগত মানগুলিকে এক এক বিশেষ ধারায় সজ্জিত না করিলে বিক্ষিপ্ত মান আমাদের চক্ষুর পীড়াদায়ক হইয়া উঠে। ঐ বিক্ষিপ্ত মানের সারাংশ করা প্রয়োজন। কারণ তাহা না হইলে তাহাদের বিশ্লেষণ, বিচার, সমালোচনা বা ব্যাখ্যা করা অসম্ভব হইয়া উঠে। তাই পরিসংখ্যান পদ্ধতি ঐ মান গুলিকে শ্রেণীবদ্ধ করিতে এবং এক নজরে সংক্ষিপ্তসার প্রদানে সাহায্য করে।

(খ) ১৯৬০ খ্রীঃ পশ্চিমবংগে প্রায় সোওয়া লক্ষ ছাত্র-ছাত্রী স্কুল ফাইনাল পরীক্ষা দিয়াছে। প্রতিবারই কমবেশী ছাত্র-ছাত্রী পরীক্ষা দিতেছে বা দিবে। পরীক্ষায় উত্তীর্ণ ছাত্র-ছাত্রীর শতকরা হিসাব দেখিয়া এবারের ছাত্র-ছাত্রীর যোগ্যতার standard বুঝা যাইতে পারে না। গত পাঁচ বৎসর যাবৎ যে Standard-এ পরীক্ষার্থী পাশ করিয়াছে এ বৎসর কি সেই Standard বজায় আছে—বাড়িয়াছে—না কমিয়াছে? তাই প্রতি বৎসরের ছাত্র-ছাত্রীর গড় যোগ্যতা বাহির করা দরকার। সেই গড় হইবে সেই বৎসরের পরীক্ষার্থীদের যোগ্যতার প্রতিভূ। বৎসর বৎসর এই যোগ্যতার প্রতিভূদের তুলনামূলক

বিচার চলিতে পারে। ঠিক এই ভাবে বিভিন্ন বিদ্যালয়ের যোগ্যতার তুলনাও সম্ভব।

(গ) অনেক সময় কেবলমাত্র প্রতিভূ সমগ্রদলের বিশেষ বিবরণ দিতে পারে না। দুই দলের প্রতিভূর এক মান হইতে পারে; কিন্তু হয়ত প্রথম দলের বিস্তার দ্বিতীয় দলের বিস্তার অপেক্ষা অনেক বেশী। প্রথম দলে পাঁচমেশালী এবং দ্বিতীয়দলে প্রায় সমযোগ্যতা বিশিষ্ট মান বিদ্যমান থাকায় ইহা সম্ভব হইতে পারে। তাই বিস্তৃতির পরিমাপ দ্বারা দলের সামগ্রিক রূপ পাওয়ার সম্ভাবনা থাকে।

(ঘ) অনেক সময় স্থায়ীভাবে সংরক্ষণ, বিভিন্ন মানের তুলনা এবং সাধারণের বোধগম্য করার জন্য মানগুলিকে চিত্রের সাহায্যে প্রকাশ করা হয়। তাই পরিসংখ্যান পদ্ধতি লেখচিত্রের সাহায্যে ঐগুলির রূপদানে সক্ষম হয়।

(ঙ) পাঠ্যক্রমে তালিকাভুক্ত বিষয়গুলির মধ্যে সম্পর্ক নির্ণয়ে পরিসংখ্যান পদ্ধতি এক বিশেষ ভূমিকায় অংশ গ্রহণ করে। তাহা ছাড়া নূতন কোন বিষয়কে পাঠ্যক্রমে অন্তর্ভুক্ত করা, বিদ্যালয়ে ছাত্র-ছাত্রী ভর্তির সময় বা কোন বৃত্তিনিরূপক পরীক্ষার বিষয়বস্তু নির্ধারণে ও প্রয়োগে অনুবন্ধ-পদ্ধতি বিশেষভাবে সাহায্য করে। কোন এক বিষয়ের ফলের উপর নির্ভর করিয়া সম্পর্কযুক্ত অন্য ফলের আভাস প্রদান এই পদ্ধতিতে সম্ভবপর।

(চ) পূর্বেই বলা হইয়াছে যে পরিসংখ্যান পদ্ধতি ভবিষ্যৎ বাণী করিতে সক্ষম। কিন্তু সম্ভাবনার নিয়মের উপর সেই ভবিষ্যৎবাণী নির্ভরশীল। তাই পরিসংখ্যানকে Normal Probability curve-এর সাহায্য লইতে হইয়াছে। এই পদ্ধতির সাহায্যে পরীক্ষার্থীদের শ্রেণীবিভাগ, বিভিন্ন সীমার মধ্যে অবস্থিত শতকরা হারে তাহাদের সংখ্যা নির্ণয়—বিভিন্ন মানের ছাত্র-ছাত্রীর তুলনা—নির্দিষ্ট বয়সের জন্য অভীক্ষা নির্বাচন প্রভৃতি অত্যাবশ্যকীয় কার্যে এই পদ্ধতি অপরিহার্য।

উল্লিখিত বিষয় ছাড়াও শিক্ষা ও মনস্তত্ত্বের অন্যান্য বিষয়ে পরিসংখ্যান পদ্ধতি বিশেষ প্রয়োজনীয়। আমাদের আলোচনার গণ্ডী সীমিত হইলেও ইহার দ্বারা বিদ্যালয়ে নিত্য প্রয়োজনীয় সমস্যাগুলিকে সাংখ্যমানের সাহায্যে রূপ দিয়া সাধারণের উপযোগী করিয়া তোলা হইয়াছে।

একথা ঠিকই যে পরিসংখ্যান পদ্ধতিতে যাদু নাই—যে বিদ্যালয়ে পাঠদান বা পরিচালনার অধিকাংশ সমস্যা এক মন্ত্রে অপসারিত হইবে। তবে এটা ঠিক যে শিক্ষক-শিক্ষিকা যদি সজ্ঞানে এই পদ্ধতির কিছুটা অনুসরণ করেন তবে

সমস্যার কিয়দংশের যে সমাধান মিলিবে তাহাতে সন্দেহ নাই। পরিসংখ্যান শিক্ষাকে পরিচালিত করিতে পারে না সত্য তবে পরিসংখ্যান যাচাই করিতে পারে—যে যতটা শিক্ষা দেওয়া হইতেছে বা যে সকল উদ্দেশ্যে শিক্ষা দেওয়া হইতেছে তাহার কতখানি সফল হইতেছে। তাই পরিসংখ্যান পরীক্ষা নহে—ইহা পরীক্ষার পরীক্ষা। শিক্ষার উদ্দেশ্য আছে—সেই উদ্দেশ্য শিশুর সামগ্রিক পরিবর্তন আনয়ন করে। এই পরিবর্তন আসে শিক্ষাগত অভিজ্ঞতার মধ্য দিয়া। কর্ম এই অভিজ্ঞতা প্রদানে সক্ষম। কিন্তু এইখানেই শিক্ষার শেষ নয়। তাহার মূল্যায়ন করা দরকার। মূল্যায়ন না করিলে শিক্ষার উদ্দেশ্য সফল হয় না। মূল্যায়নের মাধ্যমে শিক্ষাচক্র সম্পূর্ণ হয়। সেই মূল্যায়নের হাতিয়ার হইল—পরিসংখ্যান পদ্ধতি।

সেই আদর্শ সম্মুখে রাখিয়া পরিসংখ্যান পদ্ধতিকে শিক্ষাবিদ্দের হস্তে সমর্পণ করা হইয়াছে। নিত্য নূতন প্রয়োগের মাধ্যমে সকল রকম ব্যর্থতা অপসারিত হউক এবং বিপরীত ক্রমে প্রয়োগসিদ্ধ হইয়া পরিসংখ্যান নিত্য নূতন পদ্ধতি সৃষ্টি করুক ও ধীরে ধীরে পরিপূর্ণতার পথে অগ্রসর হউক !

প্রথম অধ্যায়

Frequency Distribution

যতগুলি পদ্ধতি পরিসংখ্যানের dataগুলিকে সুসজ্জিত ও সারার্থ নির্ণয়ে সাহায্য করে, Frequency distribution তাহাদের মধ্যে একটি। এই পদ্ধতিতে নম্বর বা মানকে বিভিন্ন ভাগে এবং ভাগগুলিকে বিভিন্ন গুচ্ছে সজ্জিত করা হয়। Data কখনও শ্রেণীবদ্ধভাবে (grouped) আবার কখনও বা শ্রেণীবিহীনভাবে (ungrouped) শিক্ষাবিদের কাছে আসে। প্রথমে আমরা শ্রেণীবিহীন কাঁচা (raw) dataকে কি করিয়া শ্রেণীবদ্ধ করা যায় তৎসম্বন্ধে আলোচনা করিব।

অভীক্ষার সাহায্যে কোন ব্যক্তি বিশেষের ফলকে আমরা Score বা নম্বর আখ্যা দিয়া থাকি। সাধারণতঃ এক ব্যক্তি একটি নির্দিষ্ট সময়ে বিষয়ের কতখানি নির্ভুলভাবে উত্তর করিল বা বিষয়টি কতখানি সময়ে শেষ করিল তাহা বিচার করিয়া আমরা তাহাকে নম্বর দিয়া থাকি।

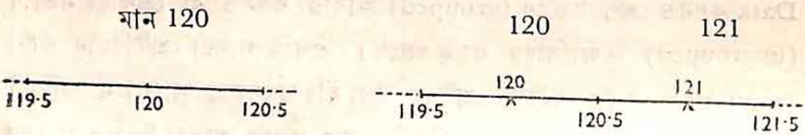
বহু মানসিক অভীক্ষা সমান এককমান দ্বারা বিভক্ত নয়। আবার বহু প্রখ্যাত মনোবিজ্ঞানী অভীক্ষায় সমান একক মান ব্যবহার করেন। কিন্তু শূন্যমান দেওয়ার পক্ষপাতী তাঁরা নন। কারণ দুজনের অভীক্ষার ফল শূন্য হইলে তো তাহাদের জ্ঞানের সীমা কোথায় বা এক ব্যক্তি অপেক্ষা অপর ব্যক্তির পার্থক্য কতখানি তাহা না ধরা পড়ার দরুণ ঐ অভীক্ষা অচল হইবে। তেমনই সম্ভবপর নয় সর্ব্বোচ্চ মান পাওয়া—যেমন 100 এর মধ্যে 100 পাওয়া।

মানসিক ও প্রাকৃতিক গুণ সমূহের নির্ণীত বিভিন্ন মানগুলি যতিবিহীন (cotinuous)। যে সারিতে যে কোন মান থাকা সম্ভবপর তাহাকে যতিবিহীন সারি বলে। যেমন ধরা যাউক কোন ছাত্রের I. Q. 120·5 ; এমন ছাত্র পাওয়া সম্ভবপর যাহার I.Q. 120·57 বা 120·568 ; মানসিক ও প্রাকৃতিক গুণসমূহের মান এই শ্রেণীভুক্ত।

আবার এমন সারি পাওয়া সম্ভব যাহার মাঝে ফাঁকা স্থান আছে। এই সকল সারির নাম যতিবহুল (discrete) সারি। যেমন পরিসংখ্যানে অঙ্ক কষিয়া বাহির হইল যে কোন গ্রামে প্রতি পরিবারে আট বৎসর বয়স্ক গড়ে 2·72 জন শিশু আছে। কিন্তু 2·72 জন শিশু কি করিয়া সম্ভবপর? হয় দুইজন নতুবা তিনজন হইবে—শিশুর ভগ্নাংশ অসম্ভব। তাই লোকসংখ্যা প্রভৃতি মানকে যতিবহুল সারির অন্তর্ভূক্ত করা হইয়া থাকে। সৌভাগ্যের

বিষয় এই যে শিক্ষা এবং মনস্তত্ত্বের মানসমূহ হয় যতিবিহীন নতুবা যতিবিহীন ভাবে ধরা হইয়া থাকে।

কাঁচা নম্বর :—এখন যতিবিহীন সারিতে একটি নম্বরের কি অর্থ? এখানে নম্বরকে একটি সম্পর্কশূন্য যতিবহুল মান ধরিলে ভুল হইবে। দুইটি সীমার মাঝে তার অবস্থিতি। যেমন 120 বলিলে 119·5 এবং 120·5 এর দূরত্বকে বুঝায়। এখানে 120 এর নিম্ন সীমা (lower limit) 119·5 এবং উচ্চ সীমা (upper limit) 120·5।



চিত্র নং 1

Array :—যখন কাঁচানম্বরগুলি শ্রেণীবিহীন অবস্থায় থাকে তখন তাহাদের সম্বন্ধে সাধারণ গুণগুলি ধরা খুবই অসুবিধাজনক। এমন কি সংখ্যা বেশী হইলে উচ্চতম ও নিম্নতম মান শীঘ্র বাহির করাও সময় ও কষ্টসাপেক্ষ। নিম্নের উদাহরণে তাহা প্রতীয়মান।

79·6	72·5	71·2	84·9
81·6	66·7	74·5	90·2
75·5	75·9	82·6	69·5
79·5	80·6	79·8	73·7
68·3	90·5	68·5	80·1

1 নং টেবল

আবার উপরোক্ত মানগুলির মধ্যে কোন্ মানের দিকে উহাদের সাধারণ ঝোঁক বা ঝোঁক মোটের উপর আছে কিনা—তাহা নির্ণয় করা সহজসাধ্য নয়। উপরের মানগুলি মানের ক্রম অনুসারে সজ্জিত হইলে নিম্নরূপ আকার ধারণ করে।

90·5	80·6	75·9	71·2
90·2	80·1	75·5	69·5
84·9	79·8	74·5	68·5
82·6	79·6	73·7	68·3
81·6	79·5	72·5	66·7

2 নং টেবল

উল্লিখিত মানসমূহ হইতে সর্ব্বোচ্চ ও সর্ব্বনিম্ন মানদুইটি বিনা আয়াসেই নির্দেশ করা যায় এবং কোন্ পর্য্যায়ে কয়টি নম্বর রহিয়াছে তাহাও নির্ধারণ করা যায়। এক কথায় 1 নং টেব্‌ল অপেক্ষা 2 নং টেব্‌ল অধিক মূল্যবান ও কার্যকরী। এইরকমভাবে সজ্জিত মানগুলির নাম Array. প্রথমতঃ ইহাতে মানগুলি ক্রমান্বয়ে সাজান থাকে। তাহা সর্ব্বোচ্চ ও সর্ব্বনিম্ন মান নির্দ্দেশক এবং মানগুলি যে 66·7 এবং 90·5 এর মধ্যে আছে তাহা স্থির। দ্বিতীয়তঃ 81 থেকে 70 এর মধ্যে যে বেশী সংখ্যক মান আছে Array তে তাহাও প্রতীয়মান। তৃতীয়তঃ উল্লিখিত মানসমূহকে কয়েকটি গুচ্ছে বিভাজন সম্ভবপর। চতুর্থতঃ উপরোক্ত মানসমূহের মধ্যে একটি যতিহীন ধারা (degree of continuity) বিদ্যমান।

কিন্তু Array তৈরী করা যেমন সময় সাপেক্ষ—তেমনই এর বিস্তৃতি। তাই Arrayতে সাজ্জিত না করিয়া Frequency distribution এ সজ্জিত করার প্রয়োজন অনুভূত হয়।

Frequency Distribution :—যদি মানগুলিকে শ্রেণীভুক্ত বা তাহাদের সারাংশ নির্ণয় না করা হয়, তবে অভীক্ষার সাহায্যে পরীক্ষা নিরীক্ষা মূল্যহীন হইয়া পড়ে। তাই প্রথমে গোটা নম্বরগুলিকে কয়েকটি ভাগে বিভক্ত করিতে হয় এবং পরে সেগুলিকে উপভাগে বিভক্ত করিয়া নম্বরগুলিকে শ্রেণীভুক্ত করা হয়।

কাঁচা নম্বরগুলিকে কিরূপে শ্রেণীভুক্ত করিয়া Frequency distributionএ সাজান হয় নিম্নে তাহার নিয়মাবলী প্রদত্ত হইল।

(১) প্রথমে সর্ব্বোচ্চ ও সর্ব্বনিম্ন নম্বর দুইটি বাছিয়া তাহাদের দূরত্ব নিরূপণের জন্য বিয়োগ করা হয়। এ দূরত্বকে Range বলা হয়। যেমন 3নং টেব্‌লে সর্ব্বোচ্চ নম্বর 86 এবং সর্ব্বনিম্ন নম্বর 31 ; সুতরাং Range 55 (অর্থাৎ 86—31).

(২) এখন কত ধাপে ও কি রকম চওড়া ধাপে ঐ নম্বরগুলিকে শ্রেণীভুক্ত করিতে হইবে তাহা নির্ণয় করিতে হইবে। সাধারণতঃ 3, 5, 10 একক লইয়া ধাপের বিস্তার (interval) ঠিক করা হয়। সাধারণতঃ 5 হইতে 15 ধাপে নম্বরগুলিকে সাজান হয়। Rangeকে ধাপ-দূরত্ব (step interval) দ্বারা ভাগ করিলে ধাপ সংখ্যা পাওয়া যায়। দূরত্ব যদি 5 ধরা হয়, তবে 3নং টেবলে (55÷5) =11 পাওয়া যায়। ইহা প্রকৃত ধাপ অপেক্ষা 1 ধাপ

কম। সুতরাং এখানে 11+1=12টি ধাপে নম্বরগুলিকে শ্রেণীভুক্ত করিতে পারা যায়। এখানে মনে রাখা উচিত যে যদি ধাপ-দূরত্ব 3 বা 10 লওয়া যায়, তবে ধাপ সংখ্যা যথাক্রমে 19 এবং 6 হইবে। 19 বেশী বড় এবং 6 বেশ ছোট বলিয়া ধাপের দূরত্ব 5 ধরিয়া 12 ধাপে সাজাইলে নম্বরগুলির সারাংশ অপেক্ষাকৃত ভাল হইবে।

(৩) এখন 4 নং টেব্‌লের ন্যায় শ্রেণীবদ্ধ করিয়া প্রতিটি নম্বরের জন্য বিভিন্ন ধাপে একটি করিয়া ঢেরা (Tally Mark /) দিতে হইবে। ঐ চিহ্ন দেওয়া শেষ হইলে পরবর্ত্তী columnএ প্রতি ধাপের সরাসরি Tally Mark-এর যোগফল লিখিতে হইবে। যেমন 40 এবং 42 দুইটি নম্বর 39—45 ধাপের মধ্যে আছে। তাই ঐ ধাপ বরাবর দুইটি ঢেরা (Tally mark) দিয়া পাশের columnএ 2 লেখা হইল। column এর যোগফলগুলি 4নং টেব্‌লে লিখিত আছে। ঐ column এর প্রতি ধাপ বরাবর সংখ্যাকে frequency বা সংক্ষেপে f বলা হয়। f columnটি যোগ করিলে মোট সংখ্যা (N) বুঝায়। এখানে Σf বা N = 50 এইরূপে নম্বরগুলি শ্রেণীভুক্ত হইল।

74	65	56	66	60
36	65	59	60	69
62	57	70	54	62
64	47	45	51	62
*86	40	42	51	77
55	34	80	53	63
67	**31	47	56	67
37	76	61	58	73
45	76	61	82	72
70	50	61	68	68

(3) নং টেবল : 50 জন শিক্ষার্থীর শ্রেণীবিহীন নম্বর

এখানে লক্ষ্য রাখা দরকার যে শ্রেণীভুক্তির সময় সর্বনিম্ন নম্বর আরম্ভ করা হইয়াছে 29 হইতে যদিও শ্রেণীবিহীন অবস্থায় সর্বনিম্ন নম্বর ছিল 31 এখানে 30—35 সর্বনিম্ন ধাপ লওয়াও চলে। তাহাতে বা উপরোক্ত ধাপের (29—34) মধ্যে কোন মূলগত পার্থক্য নাই।

Frequency distributionএ সারির উচ্চ ও নিম্ন সীমা নির্ধারণ সম্পর্কে ত্রিবিধ নিয়ম প্রচলিত আছে। নিম্নে A, B, C এই তিন প্রকারে সাজান দেখান হইয়াছে। প্রথম প্রকারে (A) দেখি সর্ব নিম্ন ধাপ 29 হইতে 34। ইহার অর্থ এই যে 29 হইতে 34 বাদে 34 পর্যন্ত সকল নম্বর ঐ ধাপে থাকিবে।

(ক) নম্বর	(খ) ঢেরা	(গ) f
84 হইতে 89	/	1
79 ,, 84	//	2
74 ,, 79	////	4
69 ,, 74	/////	5
64 ,, 69	///	8
59 ,, 64	///// /////	10
54 ,, 59	///// /	6
49 ,, 54	////	4
44 ,, 49	////	4
39 ,, 44	//	2
34 ,, 39	///	3
29 ,, 34	/	1

Σf বা $N = 50$

4 নং টেবল (ক) উপরোক্ত 50 জন শিক্ষার্থীর শ্রেণীভুক্ত নম্বর।

দ্বিতীয় প্রকারে (B) দেখি সর্বনিম্ন ধাপ 28·5 হইতে 33·5। পূর্বাপেক্ষা এই প্রকারটি বেশী স্পষ্ট। পূর্বেই বলা হইয়াছে যে 29, 28·5 হইতে 29.5 পর্যন্ত বিস্তৃত। সুতরাং 29 হইতে 34 না লিখিয়া 28·5 (29 এর নিম্ন সীমা) হইতে 33·5 (33 এর উচ্চ সীমা বা 34 এর নিম্ন সীমা)—লেখাই অধিক যুক্তিযুক্ত। তৃতীয় প্রকারে (C) দেখি যে উহা A হইতে অপেক্ষাকৃত স্পষ্ট এবং B হইতে অপেক্ষাকৃত অস্পষ্টভাবে প্রকাশিত।

তবে এও ঠিক যে তাড়াতাড়ি নির্ভুলভাবে ঢেরা দিবার সময় শেষোক্ত প্রকারই (C) অধিক প্রযোজ্য। এখানে 29 হইতে 33 ধাপটির অর্থ এই যে ধাপটি 28·5 হইতে 33·5 পর্যন্ত বিস্তৃত এবং পরবর্তী ধাপটি 33·5 হইতে 38·5 পর্যন্ত বিস্তৃত। সুতরাং Distributionটি এই প্রকারে এক যতিহীন সারিতে পরিগণিত হইয়াছে।

A			B			C	
নম্বর		f	নম্বর		f	নম্বর	f
84 হইতে	89	1	83·5 হইতে	88·5	1	84—88	1
79 ,,	84	2	78·5 ,,	83·5	2	79—83	2
74 ,,	79	4	73·5 ,,	78·5	4	74—78	4
69 ,,	74	5	68·5 ,,	73·5	5	69—73	5
64 ,,	69	8	63·5 ,,	68·5	8	64—68	8
59 ,,	64	10	58·5 ,,	63·5	10	59—63	10
54 ,,	59	6	53·5 ,,	58·5	6	54—58	6
49 ,,	54	4	48·5 ,,	53·5	4	49—53	4
44 ,,	49	4	43·5 ,,	48·5	4	44—48	4
39 ,,	44	2	38·5 ,,	43·5	2	39—43	2
34 ,,	39	3	33·5 ,,	38·5	3	34—38	3
29 ,,	34	1	28·5 ,,	33·5	1	29—33	1

5নং টেবল :—Frequency distributionএ ধাপের নিম্ন ও উচ্চ সীমা নির্ধারণ পদ্ধতি।

Exercises I

1. Indicate which of the following variables fall into continuous and which into discrete series : (a) time ; (b) salaries in a large business farm ; (c) age ; (d) census data ; (e) distance travelled by train ; (f) football scores ; (g) weight ; (h) numbers of pages in 100 books ; (i) mental age ; (j) A Q.

2. Write the exact upper and lower limits of the following scores in accordance with the definition of a score mentioned in the chapter I

46	137	1	6
19	336	95	13

3. For each one of the following ranges of measurements, state your judgment of (1) the best size of class interval, (2) the score limits of the lowest class interval, (3) the exact limits of the class interval, and (4) the midpoint.

a. 83 to 197   b. 4 to 39
c. 17 to 32   d. 35 to 96
e. 0 to 188   f. −24 to + 28.
g. 0·141 to 0·205

4. In each of the following write (a) the exact lower and upper limits of the class intervals, and (b) the midpoint of each interval

45—47	162·5—167·5	63—67	0—9
1—4	80 up to 90	16—17	25—28

5. (a) Tabulate the following twenty five scores into two frequency distributions, using (1) an interval of three and (2) an interval of 5 units. Let the first interval begin with score 60.

72	75	77	67	72
81	78	65	86	73
67	82	76	76	70
83	71	63	72	72
61	67	84	69	64

(b) Tabulate the following 100 scores into three frequency distributions, using intervals of 3, 5 and 10 units. Let the first interval begin with 45.

63	78	76	58	95
78	86	80	96	94
46	78	92	86	88
82	101	102	70	50
74	65	73	72	91
103	90	87	74	83
78	75	70	84	98
86	73	85	99	93
103	90	79	81	83
87	86	93	89	76
73	84	82	71	94
95	86	90	73	75
82	76	83	63	56
89	75	81	105	73
73	83	85	74	95
92	103	81	98	110
85	86	85	77	98
80	84	72	78	71
81	81	81	73	92
90	85	96	96	72

6. Given the following list of scores, set up a frequency distribution making the wisest choice of class interval and class limits.

62	48	65	48	51
42	46	35	45	44
52	51	38	40	45
59	39	44	47	43
42	44	43	41	48
52	46	43	52	57
44	43	52	45	46
38	45	51	46	54
42	56	53	51	50
31	51	48	43	41

7. Given the following list of scores in a "nervousness" test and using a class interval of 5, set up a frequency distribution. In the first solution, begin with the lowest class interval with a score of 35. List all exact limits of class intervals and also exact midpoints. In a second solution, start the lowest class interval with a score of 33. After finishing both solutions, write out a comparison of the two distributions and defend the choice of the one as against the other.

59	48	47	64	62	57	81	83
48	65	53	60	51	63	60	77
71	57	66	47	76	57	52	57
40	53	71	61	73	70	50	59
69	67	47	60	54	81	69	76

Answers

2. 45.5 to 46·5 ; 18·5 to 19·5 ; 136·5 to 137·5.
335·5 to 336·5 ; 94·5 to 95·5 ; 12·5 to 13·5.

3.

	i	Score limits	Exact limits
a	10	80—89	79·5—89·5
b	3	3—5	2·5—5·5
c	1	17	16·5—17·5
d	5	35—39	34·5—39·5
e	20	0—19	—0·5—19·5
f	5	—25 to—21	—25·5 to —20·5
g	·005	0·140—0·144	0·1395—0·1445

4.

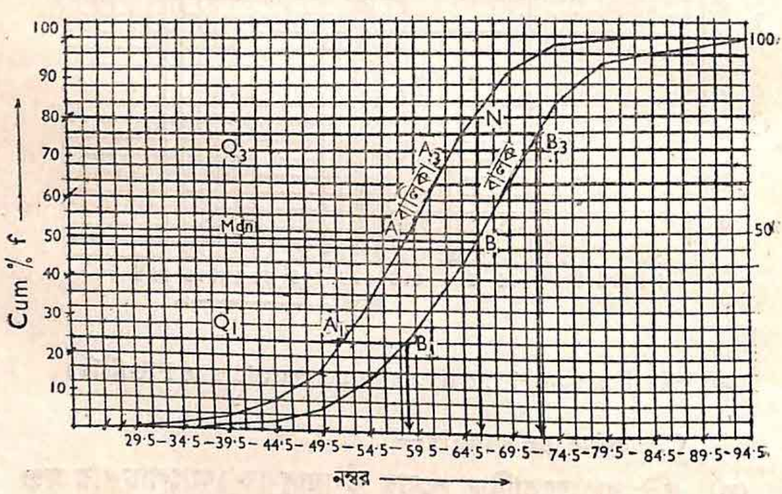
	Midpoint
44·5 to 47·5	46·0
·5 to 4·5	2·5
162·5 to 167·5	165·0
79·5 to 89·5	84·5
62·5 to 67·5	65·0
15·5 to 17·5	16·5
—·5 to 9·5	4·5
24·5 to 28·5	26·5

প্রদত্ত একটি গণিত অভীক্ষার ফলের তুলনার জন্য একই অক্ষদ্বয়ে স্থাপন করা হইয়াছে। ফলগুলি নীচে 16নং টেবলে প্রদত্ত হইয়াছে।

নম্বর (1)	বালক f (2)	cum f (3)	cum% f (4)	smoothed % f (5)	বালিকা f (6)	cum f (7)	cum% f (8)	smoothed % f (9)
90—94	0	150	100·0	100·0	0	150	100·0	100·0
85—89	1	150	100·0	99·9	0	150	100·0	100·0
80—84	7	149	99·8	98·3	1	150	100·0	99·9
75—79	15	142	95·1	93·3	0	149	99·8	99·9
70—74	28	127	85·1	82·2	6	149	99·8	98·5
65—69	37	99	66·3	64·3	25	143	95·8	91·6
60—64	25	62	41·5	44·2	40	118	79·1	75·7
55—59	18	37	24·8	27·0	31	78	52·3	54·3
50—54	10	19	12·7	14·5	25	47	31·5	32·9
45—49	6	9	6·0	6·9	11	22	14·8	17·9
40—44	3	3	2·0	2·7	7	11	7·4	8·3
35—39	0	0	0·0	·7	2	4	2·6	3·8
30—34	0	0	0·0	0	2	2	1·3	1·3
	N=150				N=150			

$$\text{Rate} = \frac{1}{150} = \cdot 0067.$$

16নং টেব্‌ল। 150 জন বালক ও 150 জন বালিকার নম্বরকে ogive লেখচিত্রে প্রকাশ করা হইয়াছে।



7নং চিত্র : 150 জন বালক ও 150 জন বালিকার উপর প্রদত্ত অভীক্ষার ফল একই অক্ষদ্বয়ে অবস্থিত দুইটি ogive-এর সাহায্যে তুলনা।

7নং চিত্রে বালক ও বালিকার ফল দুইটি ogive-এ সূচিত হইয়াছে। বহু মূল্যবান এবং চিত্তাকর্ষক তথ্য ঐ চিত্রে দেখা যায়। প্রথমতঃ বালকদের লেখচিত্র বালিকাদের লেখচিত্রের দক্ষিণ দিকে অবস্থিত। ইহার অর্থ এই যে বালকদের ফল আগাগোড়াই বালিকাদের ফল হইতে উচ্চতর। বিভিন্ন স্তরে বালক ও বালিকাদের ফলের তফাৎ ঐ লেখচিত্রদ্বয়ের মধ্যে দূরত্বের দ্বারা সূচিত হয়। দ্বিতীয়তঃ লেখচিত্রদ্বয়ের মধ্যে দূরত্বের ব্যবধান সর্বত্র সমান নহে। প্রান্ত বিন্দুদ্বয় অপেক্ষা মাঝখানে দূরত্ব বেশী। চিত্রে AB, A_1 B_1 এবং A_3B_3 প্রত্যেকটি হইতে বৃহত্তর। ইহার অর্থ এই যে মাঝখানে দুই দলের পার্থক্য বিশেষ ভাবে বিদ্যমান। দুই দলের Q_3 এবং Q_1 দ্বয়ের পার্থক্য অপেক্ষা Mdn. এর পার্থক্য বেশী তাহাও লেখচিত্রে প্রতীয়মান হয়।

আবার যদি B বিন্দুর মধ্য দিয়া লম্বটিকে উপর দিকে বর্দ্ধিত করা যায়, তবে তাহা বালিকাদের লেখচিত্রকে N বিন্দুতে ছেদ করে। N বিন্দুর দ্বারা ইহাই স্পষ্ট হয় যে বালকদের Mdn. এর নীচে প্রায় 78% বালিকার ফল বিদ্যমান; অর্থাৎ বালকদের Mdn. এর উপরে মাত্র 22% জন বালিকার ফল আছে। এইরূপে বালিকাদের Mdn. এর উপরে বা নীচে শতকরা কত বালক আছে তাহাও জানা যায়। শুধু Mdn. কেন দুই দলের যে কোন percentile-এর তুলনামূলক আলোচনা সম্ভবপর।

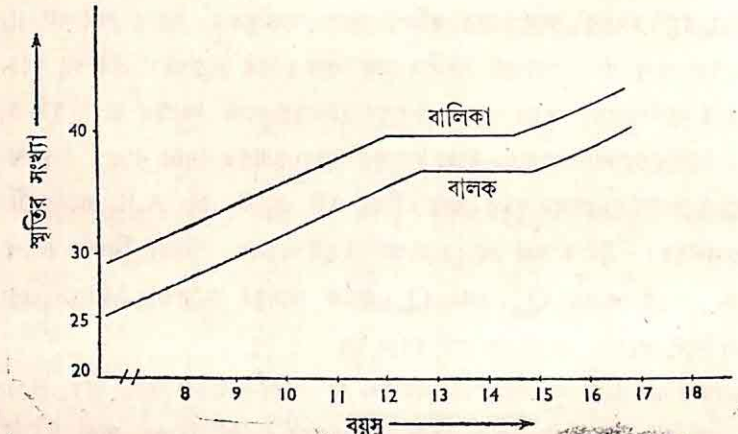
(খ) Percentile Norm নির্ণয়ে সাহায্য :

Ogive আমাদিগকে Norm নির্ণয় করিতে সাহায্য করে। শিক্ষাক্ষেত্রে Norm-এর মূল্য খুব বেশী। কোন শিক্ষাগত অভীক্ষার Norm বাহির করা থাকিলে তাহার সাহায্যে অন্য কোন ছাত্রের (সেই অভীক্ষার) ফলকে তুলনা করা হইয়া থাকে। Norm না বাহির করিলে অভীক্ষা প্রস্তুতির অর্থ মূল্যহীন হইয়া পড়ে। সাধারণতঃ বড় দলের উপর প্রয়োগসিদ্ধ Mean বা Mdn কে Norm ধরা হইয়া থাকে।

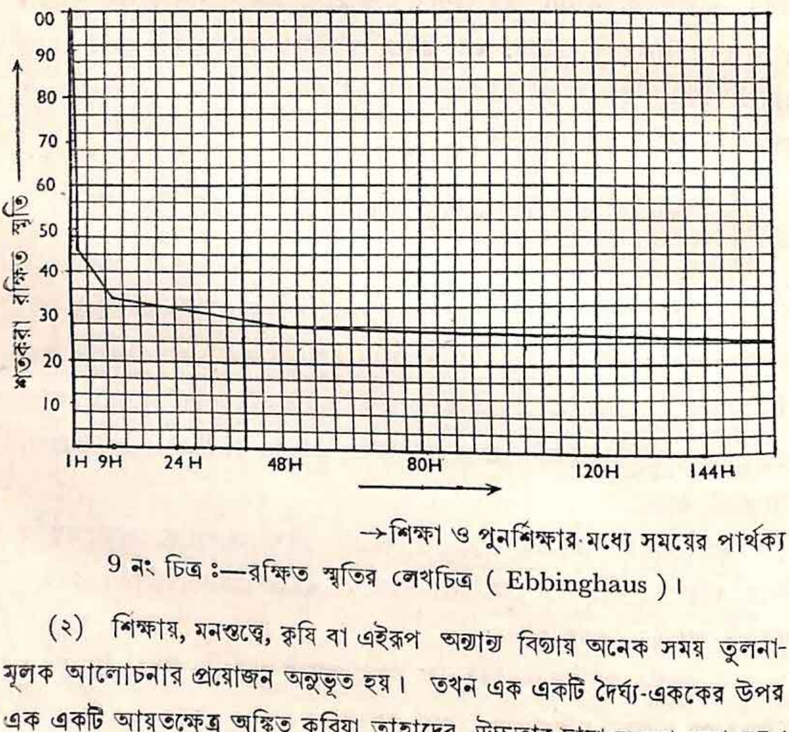
(D) উল্লিখিত লেখচিত্র ছাড়া বৃদ্ধি, অভ্যাস, শিক্ষা, ব্যবসায়, কৃষি প্রভৃতি বিষয়ে পরিবর্ত্তন নির্দেশক লেখচিত্র আছে। নীচে উদাহরণ স্বরূপ এইরূপ কয়েকটি লেখচিত্র দেওয়া হইল।

(১) লাইন লেখচিত্র :—(ক) 8নং চিত্রে বালক ও বালিকাদের বয়োবৃদ্ধির সংগে সংগে যুক্তিপূর্ণ স্মৃতির বৃদ্ধি দেখান হইয়াছে।

(খ) 9 নং চিত্রে বিখ্যাত মনোবিজ্ঞানী Ebbinghaus এর সময়ের ব্যবধানে রক্ষিত স্মৃতির লেখচিত্র দেওয়া হইল।



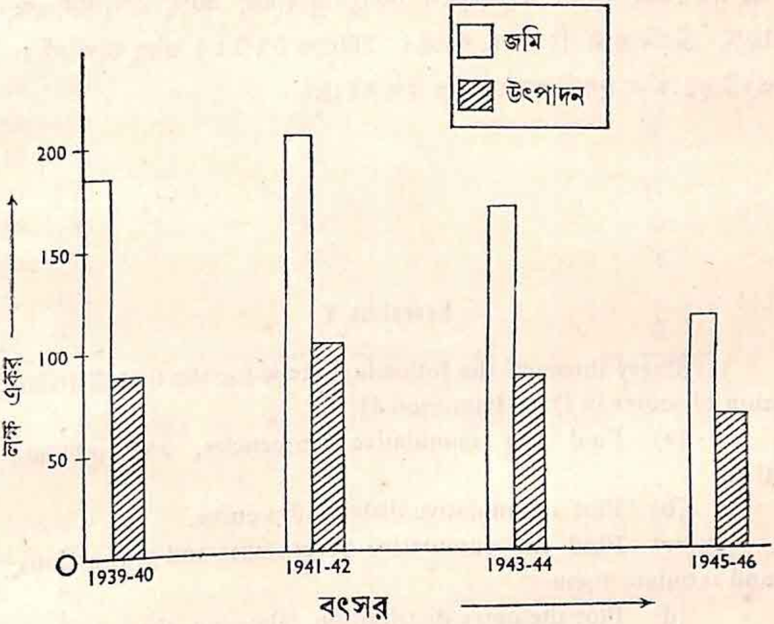
৪নং চিত্র ঃ বয়সের সংগে সংগে যুক্তিপূর্ণ স্মৃতির বৃদ্ধি (Garrett).



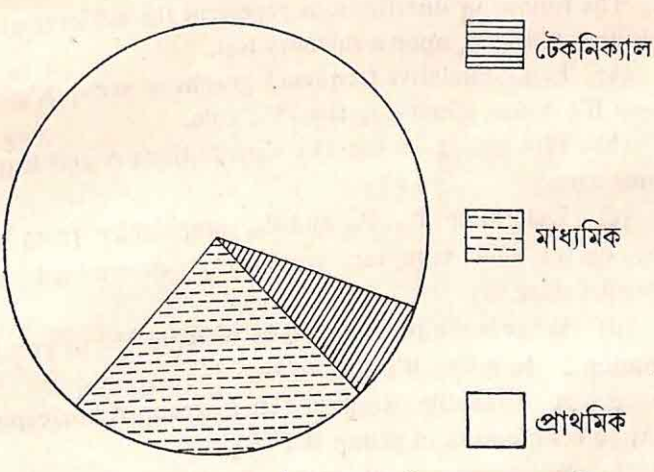
9 নং চিত্র ঃ—রক্ষিত স্মৃতির লেখচিত্র (Ebbinghaus)।

(২) শিক্ষায়, মনস্তত্ত্বে, কৃষি বা এইরূপ অন্যান্য বিদ্যায় অনেক সময় তুলনামূলক আলোচনার প্রয়োজন অনুভূত হয়। তখন এক একটি দৈর্ঘ্য-এককের উপর এক একটি আয়তক্ষেত্র অঙ্কিত করিয়া তাহাদের উচ্চতার দ্বারা তুলনা করা হয়।

নীচে 10 নং চিত্রে অবিভক্ত ভারতের বিভিন্ন সালে তূলার জমি ও উৎপাদনের হার লেখচিত্রের দ্বারা প্রকাশ করা হইল।



10 নং চিত্র : অবিভক্ত ভারতে তূলার জমি ও উৎপাদনের লেখচিত্র।



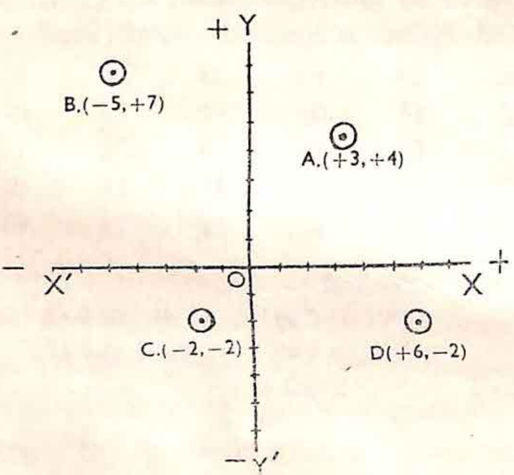
11নং চিত্র : ভারতের একটি শিল্প অধ্যুষিত অঞ্চলে প্রাথমিক, মাধ্যমিক ও টেকনিক্যাল বিদ্যালয়ের অনুপাত।

দ্বিতীয় অধ্যায়

Frequency distribution-কে লেখচিত্রে প্রকাশ

সাংখ্যমানকে (Numerical data) লেখচিত্রে প্রকাশ করিলে বিশ্লেষণের সুবিধা হয়। কোন distribution-কে লেখচিত্রে রূপ দিলে তাহা চক্ষুর সম্মুখে জীবন্ত বলিয়া প্রতিভাত হয়। শুধু সংখ্যা যেখানে চক্ষুর পীড়াদায়ক হইয়া উঠে সেখানে তাহার লেখচিত্র চক্ষু ও মনের উপাদেয় খোরাক হইয়া দাঁড়ায়। সাধারণতঃ চারি প্রকারে এই লেখচিত্রের প্রকাশ দেখা যায়—(ক) Frequency Polygon, (খ) Histogram, (গ) Cumulative frequency curve এবং (ঘ) Cumulative percentage curve বা ogive. প্রথম দুইটির আলোচনা এই অধ্যায়ে এবং (গ) ও (ঘ) দুইটি পঞ্চম অধ্যায়ে আলোচিত হইবে।

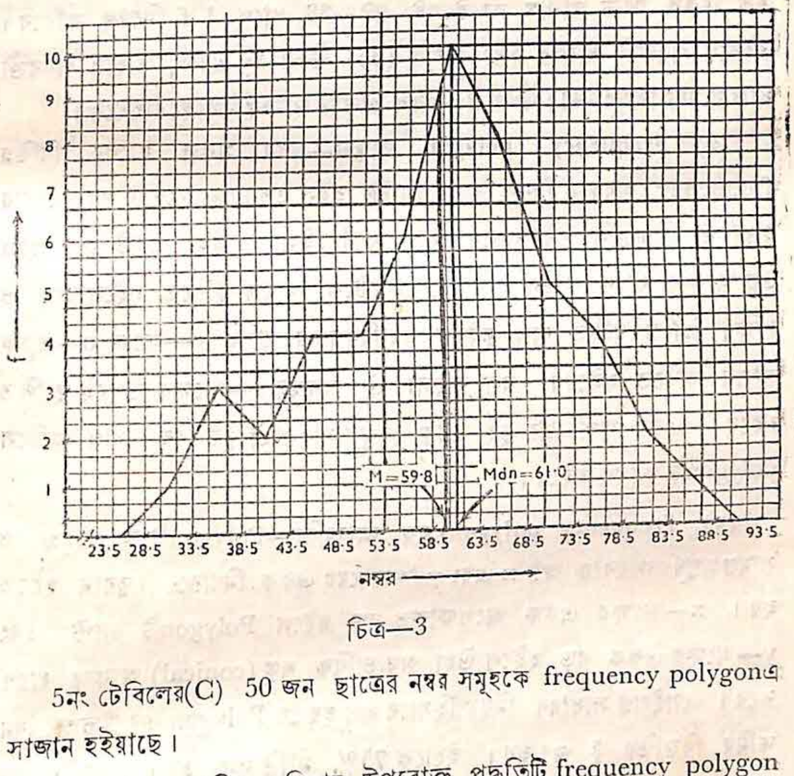
নিম্নের চিত্রে দেখা যাইতেছে যে দুইটি অক্ষ xox′ এবং yoy′ পরস্পর o মূল বিন্দুতে লম্বভাবে ছেদ করিয়াছে এখন x—অক্ষের উপর o বিন্দু হইতে যদি 3 একক ডান দিকে যাওয়া যায় এবং যদি ঐ দিকে যাওয়াকে '+' গতি বলা হয়, তবে ঐ স্থানের বিন্দুর অবস্থানকে +3 দ্বারা প্রকাশ করা হইবে। এইরূপে ঐ একই অক্ষের উপর যদি o হইতে 5 একক বাম-দিকে যাওয়া হয়



চিত্র—2

তবে ঐ অবস্থানকে (বিপরীত গতি বলিয়া) '—5' দ্বারা প্রকাশ করা হয়। এইরূপে y—অক্ষের উপর o মূল বিন্দু হইতে উপরে বা নীচে যাইলে যথাক্রমে '+' এবং '—' চিহ্ন দিয়া অবস্থানগুলির মার্কা করিয়া দিতে

হয়। আবার যদি (+3, +4) বিন্দুটিকে লেখচিত্রে প্রকাশ করিতে বলা হয়, তখন বুঝিতে হইবে যে মূলবিন্দু হইতে x-অক্ষ বরাবর +3, এবং সেই বিন্দু হইতে y—অক্ষ বরাবর +4 একক গিয়া যে বিন্দু পাওয়া যাইবে সেই বিন্দুর স্থানাঙ্ক (+3, +4)। এখানে চিত্রে(2) A সেই রকম একটি বিন্দু। একই প্রক্রিয়ায় B (—5, +7), C(—2, —2) এবং D (+6, —2) বিন্দুগুলি লেখচিত্রে স্থাপন করা হইয়াছে।



চিত্র—3

5নং টেবিলের(C) 50 জন ছাত্রের নম্বর সমূহকে frequency polygonএ সাজান হইয়াছে।

A. স্থানাঙ্ক বাহির করিবার উপরোক্ত পদ্ধতিটি frequency polygon তৈরীর ক্ষেত্রে বিশেষ উপযোগী। নীচে frequency polygon তৈরীর নিয়ম প্রদত্ত হইল। (5নং টেবল (C) দ্রষ্টব্য)

(ক) **ভূমির উপর নম্বর গুলিকে সাজান**—Frequency polygon-এর ভূমিতে বা x—অক্ষে নম্বরগুলিকে ধাপ অনুযায়ী সাজাইবার বিভিন্ন নিয়ম বিদ্যমান। কোন কোন সময় ইহাদিগকে চিত্র—3এর ন্যায় নম্বরগুলির সীমানা দিয়াই ভূমিতে সজ্জিত করা হয়। আবার কখন কখন টেবল অনুযায়ী

নম্বরগুলিকে ভূমির উপর স্থাপন করা হয়। এই স্থাপনার সময় দুই প্রান্তে দুইটি অতিরিক্ত ধাপ লওয়া হয়। ভূমিতে '।।' দাগদিবার অর্থ ভূমিতে নম্বরগুলিকে ইচ্ছা মত বামে বা দক্ষিণে সরান যাইতে পারে।

(খ) **f গুলিকে লেখচিত্রে স্থাপন**—প্রতি ধাপের fকে ঐ ধাপের মধ্য বিন্দুর উপর দাঁড় করাইতে হয়। সুতরাং 28·5 হইতে 33·5 পর্যন্ত ধাপের মধ্য বিন্দু হইবে $\left(\frac{28 \cdot 5 + 33 \cdot 5}{2} = \right) 31$. সুতরাং 31 নম্বরের উপর y—অক্ষে এক একক উচ্চে স্থানাঙ্ক বসাইলেই 29—33 ধাপে 1 f নির্দেশ করিবে। এইরূপে পরবর্তী ধাপের মধ্য বিন্দুর (36) উপর 3y-একক, তাহার পরবর্তী ধাপের মধ্য বিন্দুর (41) উপর 2y-একক প্রভৃতি স্থানাঙ্ক বসাইতে হইবে।

(গ) **Frequency polygon অঙ্কন**—যখন সকল ধাপের বিভিন্ন frequency অবস্থান বিন্দুর দ্বারা নির্দিষ্ট হইল তখন ছোট ছোট সরল রেখার দ্বারা ঐ বিন্দুগুলি পর্যায়ক্রমে যোগ করা হইল। ইহা একটি polygon এর আকার ধারণ করিল। কিন্তু ক্ষেত্রটিকে আবদ্ধ করিতে হইলে বাম ও দক্ষিণ প্রান্তে বর্দ্ধিত ধাপ দুইটির মধ্যবিন্দুদ্বয়ের উপর y—অক্ষে o—একক স্থাপন করিতে হইবে। বলা বাহুল্য ঐ বিন্দুদ্বয় x—অক্ষের উপরই স্থাপিত হইবে। এখন লেখচিত্রের দুই প্রান্ত বিন্দুর সহিত ঐ দুটি বিন্দু যোগ করিলে polygonটি আবদ্ধ হইবে।

(ঘ) **Frequency polygon-এর আকৃতি**—Polygonটিকে সুদৃশ্য ও সামঞ্জস্যপূর্ণ করিবার জন্য x এবং y অক্ষদ্বয়ের একক নির্ধারণে যত্নবান হইতে হয়। x—অক্ষের একক অপেক্ষাকৃত বড় হইলে Polygonটি চ্যাপ্টা এবং y—অক্ষের একক বড় হইলে উহা অস্বাভাবিক শঙ্কু (conical) আকার ধারণ করে। সেইজন্য সাধারণ নিয়ম হিসাবে ধরা হয় যে Polygon এর উচ্চতা যেন ভূমির বিস্তৃতির $\frac{3}{4}$ গুণ হয়। ইহাকে 75% নীতি বলা হয়। বলা বাহুল্য যে কার্যকালে হয়ত y—অক্ষ x—অক্ষের 75% ভাগের কিছু কম বা বেশী হয়। চিত্রে—(3) দেখা যায় যে উহাতে মোট 13টি ধাপ আছে। ছক কাগজে যদি একটি ধাপ 6টি ক্ষুদ্র বর্গ এককের দৈর্ঘ্য দ্বারা সূচিত হয়, তবে 13 ধাপের মোট দৈর্ঘ্য হইবে $13 \times 6 = 78$. এখন 75% নীতি অনুসরণ করিলে Polygon-এর উচ্চতা হইবে 78 এর 75% অর্থাৎ প্রায় 59 একক। এখন 59কে বৃহত্তম f দ্বারা ভাগ করিলে y—অক্ষে এক f-এর একক পাওয়া যাইবে। এখানে বৃহত্তম

f হইতেছে 10—সুতরাং 59 ÷ 10 = 6 প্রায়। y—অক্ষে 6টি ক্ষুদ্র বর্গক্ষেত্রের বাহুকে 1 f ধরা হইল এবং ক্রমান্বয়ে উপরে একই একক লইয়া 1,2,3,…10 লেখা হইল। পরে পূর্ব বর্ণিত নিয়ম অনুসারে প্রতি ধাপের মধ্যবিন্দুর উপর নির্দিষ্ট f স্থাপন করা হইল।

সকল Frequcncyর যোগফল অর্থাৎ N উক্ত polygon এর ক্ষেত্রফল দ্বারা সূচিত হয়। প্রতি ধাপের উপর ক্ষেত্রটি সমগ্র ক্ষেত্রের অর্থাৎ N এর আনুপাতিক মূল্য নির্ধারণ করা উচিত। তবে যেহেতু polygon টির মস্তক পর্য্যায়ক্রমে বক্রতালিক নয়—তাই কিছু পার্থক্য পরিলক্ষিত হয়।

Frequency Polygonকে Smooth করার নিয়ম :—যেহেতু মোট সংখ্যা অর্থাৎ N বেশ ছোট এবং f মধ্য ধাপের দুই পার্শ্বে অসামঞ্জস্য ভাবে সজ্জিত, তাই উপরোক্ত Polygonটি ঠিক Smoothed নয়। সংখ্যা বেশী পরিমানে হইলে Smoothed হওয়ার সম্ভাবনা থাকিত। উক্ত Polygon টিকে Smoothed করিতে হইলে f স্তম্ভকে পান্টাইয়া নূতন আনুমানিক গড় f বসাইতে হয়। নিম্নে নিয়ম প্রদত্ত হইল। 64—68 ধাপের f কে Smoothed করিলে $\frac{5+8+10}{3} =$ 7·67, এবং 44—48 ধাপের f কে Smoothed করিলে $\frac{4+4+2}{3} = 3{\cdot}33$ ইত্যাদি হইবে। অর্থাৎ যে ধাপের f কে Smothed করিতে হইবে সেই ধাপের f এবং তাহার অব্যবহিত উপর ও নীচের f দুইটির যোগফল নির্ণয় করিয়া 3 দ্বারা ভাগ করিলেই সেই ধাপের f টি Smoothed হইবে। Smoothed f গুলি নূতন টেবলে (4 নং খ) প্রদত্ত হইল। বলা বাহুল্য যে সর্বোপরি এবং সর্বনিম্ন ধাপের f কে Smoothed করিতে গেলে পূর্বের টেব্‌লে যথাক্রমে উপরে ও নীচে নূতন একটি করিয়া অতিরিক্ত ধাপ লিখিতে হয়—যদিও সেই অতিরিক্ত ধাপ দুইটির f, o হইবে।

নম্বর	f	Smoothed f
89—93	0	·33
84—88	1	1·00
79—83	2	2·33
74—78	4	3·67
69—73	5	5·67
64—68	8	7·67
59—63	10	8·00

নম্বর	f	Smoothed f
54—58	6	6·67
49—53	4	4·67
44—48	4	3·33
39—43	2	3·00
34—38	3	2·00
29—33	1	1·33
24—28	0	·33
	50	50

4 নং টেবল (খ) পূর্ব f এবং Smoothed f পাশাপাশি সজ্জিত।

উপরোক্ত Smoothed f কে দ্বিতীয় বারের জন্য আবার Smoothed করা যায় এবং এইরূপে frequency polygonকে Smoothed করা হয়। যদি N খুব বেশী হয়, তখন হয়ত Smoothed করার প্রয়োজন অনুভূত নাও হইতে পারে।

B. Histogram বা স্তম্ভ লেখচিত্র

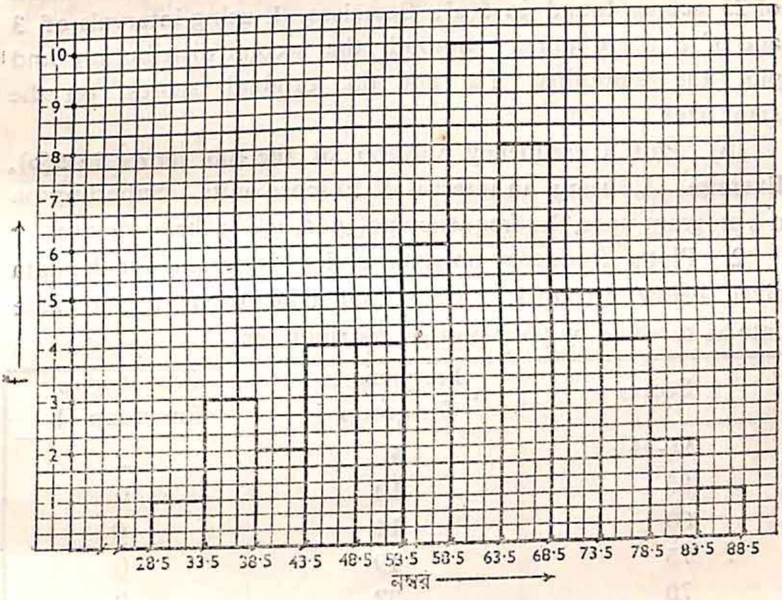
Frequency distributionকে লেখচিত্রে প্রকাশ করিবার দ্বিতীয় উপায় হইল Histogram বা স্তম্ভলেখচিত্র অঙ্কন। 5নং টেবলের (C) distributionটি চিত্র—4 এ Histogram বা স্তম্ভলেখচিত্রে প্রকাশ করা হইয়াছে। Frequency polygon বা Histogram তৈরীর নীতি একই। তবে Histogram অঙ্কনের সময় মনে রাখা উচিত যে গোটা ধাপের উপরই f স্থাপন করিতে হয়। (কিন্তু Frequency polygon এর সময়ে ধাপের মধ্য বিন্দুর উপর f স্থাপন করা হয়) প্রতিটি ধাপের উপর এক একটি আয়ত ক্ষেত্র দ্বারা f এর মান সূচিত হয়।

নীচে চিত্র—4 এ Histogrgm এর ভূমিটি 28·5, 33·5, 38·5 প্রভৃতি দ্বারা সূচিত হইয়াছে। মনে রাখা দরকার যে Frequency distribution এর মত এখানেও অক্ষ ঠিক করার সময় 75% নীতি অনুসরণ করা হয়। 28·5—33·5 ধাপের মধ্যে f = 1 কে ঐ ধাপের দূরত্ব ও y– অক্ষে 1—একক বিশিষ্ট আয়তক্ষেত্রের ক্ষেত্রফল দ্বারা সূচিত করা হইয়াছে।

উপরের চিত্রে (4) স্পষ্টই দেখা যাইতেছে যে অনেকগুলি আয়তক্ষেত্রের উচ্চতা একই। এমন কি পাশাপাশি ধাপ দুইটির (43·5—48·5 এবং 48·5—53·5) উচ্চতা এক।

যেহেতু সব ধাপের উপরে অংকিত আয়তক্ষেত্রের দ্বারা বিভিন্ন f সূচিত

হইতেছে—সুতরাং চিত্রে প্রদত্ত ক্ষেত্রের মাঝের লম্বরেখাগুলি টানার প্রয়োজন নাই। বাইরের রেখা দ্বারাই বিভিন্ন ধাপে f এর উত্থান পতন নির্দেশিত হইতে পারে। Frequency distribution এর মত N ঐ সীমাবদ্ধ ক্ষেত্রের দ্বারা সূচিত হইয়া থাকে। প্রতিটি ধাপের f ঐ ধাপের উপর আয়তক্ষেত্রের সহিত নির্দিষ্ট অনুপাতে বিদ্যমান।



চিত্র—4

Frequency polygon এবং Histogram এর যথাযথ ব্যবহার :—কখন polygon বা Histogram ব্যবহার করা যুক্তিযুক্ত তাহার নির্দিষ্ট কোন সাধারণ নিয়ম নাই যে সর্বক্ষেত্রে প্রযুক্ত হইবে, তবে Histogram, polygon অপেক্ষা অনেক সঠিক সংবাদ পরিবেশন করে। কারণ পূর্বেই বলা হইয়াছে যে সেখানে ধাপের উপর অংকিত আয়তক্ষেত্র সেই ধাপের f এর সহিত আনুপাতিক ভাবে বিদ্যমান, polygon ঠিক তাহা নহে।

তবে এটা ঠিকই যে দুই লেখচিত্রই একই তথ্য পরিবেশন করে। কোন্ ধাপে কতটা f বা লেখচিত্রের কোন্‌দিকে কিরূপ ঝোঁক তাহা উক্ত লেখচিত্রদ্বয়ে পাওয়া যায়। প্রশ্ন ছাত্র-মেধার তুলনায় কঠিন হইলে লেখচিত্রের বাম দিক

বেশী ভারী হয় এবং বিপরীতক্রমে, ডানদিক বেশী ভারী হয়। প্রশ্ন ছাত্রোপযোগী হইলে লেখচিত্রে ধরা পড়িবে যে চিত্রের বামে বা ডাইনে কম এবং মাঝখানে বেশীসংখ্যক পরীক্ষার্থী আছে।

Exercises II

1. (a) Plot frequency polygons for the two distribution of 25 scores found in 5(a), Exercises—I, using intervals of 3 and of 5 score units. Smooth the second distribution and plot the smoothed f's and the original scores, on the same axes.

(b) Plot a frequency polygon of the 100 scores in 5(b), Exercise —I, using an interval of 10 score units. Superimpose a histogram upon the frequency polygon, using the same axes.

2, Plot a frequency polygon and a histogram for the data given below, Group 1. State your conclusions about these data as revealed by your plotted distributions.

Scores	Frequencies for group I	Frequencies for group II
90-94	4	2
85	10	0
80	14	0
75	19	0
70	32	2
65	31	4
60	40	5
55	28	12
50	29	13
45	21	21
40	18	21
35	10	19
30	6	20
25	1	14
20	3	1
Sums	266	134

3. Apply the smoothing process described in this chapter to the data given in Q. 2, Exercises—II, (Group I). Plot a curve based upon the smoothed frequencies.

4. Reduce the distributions A and B below to percentage frequencies and plot them as frequency polygons on the same axes. Is your understanding of the achievement of these groups advanced by this treatment of the data ?

Scores	Group A	Group B.
52—55	1	8
48—51	0	5
44—47	5	12
40—43	10	58
36—39	20	40
32—35	12	22
28—31	8	10
24—27	2	15
20—23	3	5
16—19	4	0
	65	175

তৃতীয় অধ্যায়

Central Tendencyর পরিমাপ

কি শিক্ষাগত কি মানসিক—কি প্রাকৃতিক—সকল রকম ক্ষমতার পরিমাপ করিবার সময় আমরা বেশী সংখ্যক f কে শ্রেণীবদ্ধ সারির মাঝামাঝি স্থানে অবস্থিত থাকিতে দেখি। সহজেই নজরে পড়ে যে মৌমাছির ঝাঁকের ন্যায় মধ্যস্থলটি স্ফীত হইয়াছে। ইহার অর্থ এই যে খুব কম বা খুব বেশী মান পাইয়াছে খুব কম সংখ্যক ছাত্র এবং মাঝামাঝি মান আছে বেশী সংখ্যক ছাত্রের। সংখ্যা বেশী হইলে এই মাঝের দিকের ঝোঁক বিশেষ ভাবে প্রতীয়মান হয়। এই ঝোঁকের নাম central tendency.

সাধারণতঃ central tendency-কে পরিমানগত ভাবে তিন প্রকারে প্রকাশ করা হয়—(ক) Mean, (খ) Median এবং (গ) Mode.

কাঁচা নম্বরগুলিকে frequency distribution-এ সজ্জিত করিলে গোটা দলের একটা সারাংশ মিলে বটে; কিন্তু তাহাতে তাহাদের প্রতিভূ হিসাবে কাহাকে ধরা যাইবে তাহা পাওয়া সম্ভব নয়। তাই দলের প্রতিভূ বাহির করা ও বিভিন্ন দলের প্রতিভূদের মধ্যে তুলনা করিতে হইলে Mean, Median বা Mode-এর দ্বারস্থ হইতে হইবে। পরিসংখ্যানে এই তিন প্রতিভূ 'গড়' নামে অভিহিত হইয়া থাকে। নিম্নে এই তিনটি প্রতিভূর পরিমাপ পদ্ধতি সম্বন্ধে আলোচনা করা হইয়াছে।

Central tendency'র পরিমাপ নির্ণয়

(ক) Mean বা M

(১) যখন নম্বরগুলি শ্রেণীবিহীন ভাবে থাকে তখন নম্বরগুলির যোগফলকে মোট সংখ্যা দ্বারা ভাগ করিলে Mean পাওয়া যায়। কোন লোক যদি প্রথম সপ্তাহে 40 টাকা, দ্বিতীয় সপ্তাহে 35 টাকা, তৃতীয় সপ্তাহে 29 টাকা এবং চতুর্থ সপ্তাহে 42 টাকা উপায় করে—তবে তাহার সাপ্তাহিক আয়ের Mean হইবে 36·50 টাকা। তাই শ্রেণীবিহীন অবস্থায় M-এর সূত্র হইবে—

$$M = \frac{\Sigma X}{N} \quad \dots\dots\dots\dots\dots\dots \quad (1)$$

এখানে মোট সংখ্যা হইতেছে N এবং নম্বরগুলির যোগফল ΣX দ্বারা সূচিত

হইতেছে। Σ চিহ্নটি যোগফলের চিহ্ন, ΣX বলিলে যত X আছে তাহাদের যোগফলকে বুঝায়।

(২) নম্বরগুলি যখন শ্রেণীবদ্ধ অবস্থায় থাকে তখন উপরোক্ত উপায়ে Mean বাহির করা সম্ভব নয়। তখন ভিন্ন প্রক্রিয়ায় Mean বাহির করিতে হয়। নিম্নে উদাহরণে 42 জন ছাত্রের Mean বাহির করা হইয়াছে। প্রথমে প্রতি ধাপের মধ্য-বিন্দু X বাহির করা হইল। এখন প্রতি ধাপের মধ্যবিন্দুকে সেই ধাপের f দ্বারা গুণ করিয়া fX সারি পাওয়া গেল। এখন ΣfX (4512) বাহির করিয়া N (42) দ্বারা ভাগ করিলেই Mean (107·43) পাওয়া যাইবে। মনে রাখা উচিত যে শ্রেণীবদ্ধ করিলে নম্বরগুলি তাহাদের বৈশিষ্ট্য হারায় এবং মধ্য বিন্দুই গোটা ধাপের প্রতিভূ হইয়া দাঁড়ায়। তাই মধ্যবিন্দুকে f দ্বারা গুণ করিলে ঐ ধাপের মধ্যে অবস্থিত নম্বরগুলির যোগফলই বুঝায়। যদি X মধ্যবিন্দু হয়, f ধাপের frequency এবং N মোট সংখ্যা নির্দেশ করে, তবে শ্রেণীবদ্ধ সারির Mean এর সূত্র হইবে—

$$M=\frac{\Sigma fX}{N} \quad \cdots\cdots\cdots\cdots\cdots\cdots \cdots(2)$$

নম্বর	মধ্যবিন্দু (X)	f	fX
120—122	121	1	121
117—119	118	3	354
114—116	115	4	460
111—113	112	7	784
108—110	109	9	981
105—107	106	4	424
102—104	103	5	515
99—101	100	4	400
96—98	97	2	194
93—95	94	2	188
90—92	91	1	91
		N=42	ΣfX=4512

$$\text{Mean}=\frac{\Sigma fX}{N}=\frac{4512}{42}=107{\cdot}43.$$

6নং টেব্‌ল : মধ্যবিন্দুর সাহায্যে শ্রেণীবদ্ধ নম্বরের Mean নির্ণয়।

(৩) যদি কতকগুলি দলের Mean এবং তাহার লোকসংখ্যা দেওয়া থাকে তবে তাহাদের Mean সমষ্টিগতভাবে বাহির করা যায়। যেমন 10 জন লোক

বিশিষ্ট একটি দলের Mean আয় 150 টাকা এবং 40 জন লোক বিশিষ্ট অন্য একটি দলের Mean আয় 140 টাকা। সমষ্টিগতভাবে তাহাদের আয়ের Mean হইবে $\frac{10\times150+40\times140}{40+10}$ অথবা 142 টাকা। সমষ্টিগত ভাবে x দলের Mean এর সূত্র হইবে

$$M_{comb.}=\frac{N_1M_1+N_2M_2+\cdots\cdots+N_xM_x}{N_1+N_2+\cdots\cdots+N_x}\quad\cdots\cdots(3)$$

যখন দুইটি দল হইবে তখন Mean হইবে

$$M_{comb}=\frac{N_1M_1+N_2M_2}{N_1+N_2}$$

(৪) শ্রেণীবদ্ধ frequency distribution এর mean বাহির করিবার সহজ উপায়—

6নং টেব্‌লে মধ্যবিন্দুকে (X) সেই ধাপের f দ্বারা গুণ করিয়া তাহাদের যোগফলকে মোট সংখ্যা (N) দ্বারা ভাগ করিয়া mean (M) বাহির করা হইয়াছে। এই পদ্ধতি যদিও সঠিক উত্তর প্রদান করে, তবু একটু দীর্ঘ এবং সময় সাপেক্ষ। তাই 'Assumed mean' ধরিয়া লইয়া সংক্ষেপে mean বাহির করা হয়। ইহাকে mean বাহির করিবার সংক্ষিপ্ত পদ্ধতি বলে। Median বা Mode বাহির করিবার এরূপ সংক্ষিপ্ত পদ্ধতি নাই।

প্রথমে আন্দাজ মত একটি ধাপে mean এর অবস্থিতি ধরিয়া লওয়া হয়। এই meanকে Assumed Mean বা AM বলে। যেহেতু আন্দাজে ধরা হইয়াছে, সুতরাং কিছু ভুল তাহাতে থাকিতে পারে। তাই AMকে আবার সংশুদ্ধ করিতে হয়। কিন্তু আন্দাজ মত mean ধরার বিশেষ কোন নিয়ম নাই। সাধারণতঃ distribution এর মাঝামাঝি স্থানে, অথবা যে ধাপে সর্বাপেক্ষা বেশী f আছে সেই ধাপের মধ্যবিন্দুকে AM ধরিতে হয়। নিম্নে প্রদত্ত 7নং টেব্‌লে সর্বাপেক্ষা f বেশী আছে 30-34 ধাপে। তাই AM ঐ ধাপের মধ্যবিন্দু অর্থাৎ 32 ধরাই ভাল। এখানে মনে রাখা দরকার যে সংক্ষিপ্ত পদ্ধতিতে যে কোন ধাপের মধ্যবিন্দুকে AM লইয়া সংশুদ্ধ করিলে mean একই হইবে।

(i) প্রথমে x^1 স্তম্ভ ভর্ত্তি করিতে হয়। AM হইতে প্রতি ধাপের মধ্যবিন্দু কত ধাপ-একক দূরে অবস্থিত তাহাই x^1 দ্বারা সূচিত হয়। যেমন 35-39 এর মধ্যবিন্দু 37. AM 32 হইতে 35—39 এর মধ্যবিন্দু 37, 1 ধাপ-একক দূরে অবস্থিত। তাই x^1 স্তম্ভে 37 বরাবর 1 লেখা হয়। এইরূপে এর পরবর্তী

উপরের ধাপগুলির মধ্যবিন্দুগুলি যথাক্রমে দুই ধাপ-একক, তিন ধাপ-একক ইত্যাদি দূরে থাকায় পরবর্তী মধ্যবিন্দুগুলি বরাবর যথাক্রমে 2, 3, 4 ইত্যাদি লেখা হয়। AM সুরুবিন্দু বলিয়া তাহার পাশে O লেখা হয় এবং নিম্নের ধাপগুলির মধ্যবিন্দুবরাবর একই নিয়মে যথাক্রমে $-1, -2, -3$ ইত্যাদি লেখা হয়।

(ii) x^1 স্তম্ভ ভর্তি হইলে fx^1 স্তম্ভ তৈরী করিতে হয়। লক্ষ্য করা যায় যে AMএর উপরে fx^1 ধনাত্মক এবং নীচে ঋণাত্মক হইবে। x^1 এর চিহ্নের উপর fx^1 এর চিহ্ন নির্ভর করে।

(iii) fx^1 স্তম্ভের ধনাত্মক ও ঋণাত্মক ফলগুলি আলাদা যোগ করা হয়, এবং উক্ত ফলদুইটির বীজগণিতিয় যোগফল নির্ণয় করা হয়। নীচের টেব্‌লে ধনাত্মক ফল $+43$ এবং ধনাত্মক ফল -55. সুতরাং তাহাদের বীজগণিতিয় যোগফল -12. এখন -12কে N অর্থাৎ 50 দ্বারা ভাগ করিলে c এবং ঐ c কে ধাপ দূরত্ব দ্বারা গুণ করিলে ci পাওয়া যাইবে। এখন AM এর সহিত ci যোগ করিলে সংশুদ্ধ Mean পাওয়া যাইবে। এখানে $c = -12/50 = -0{\cdot}24$ এবং $i = 5$ $\therefore$ $ci = -0{\cdot}24 \times 5 = -1{\cdot}20$.

$\therefore$ সংশুদ্ধ Mean $= AM + ci = 32 + (-1{\cdot}20) = 30{\cdot}80$.

নম্বর	মধ্যবিন্দু	f	x^1	fx^1
55—59	57	1	5	5
50—54	52	2	4	8
45—49	47	4	3	12
40—44	42	5	2	10
35—39	37	8	1	8
30—34	32	10	0	0
25—29	27	6	−1	−6
20—24	22	4	−2	−8
15—19	17	4	−3	−12
10—14	12	2	−4	−8
5— 9	7	3	−5	−15
0— 4	2	1	−6	−6
		N = 50		$\Sigma fx^1 = -12$

$AM = 32{\cdot}00$ $\quad c = -\frac{12}{50} = -{\cdot}240$
$ci = -1{\cdot}20$ $\quad i = 5$
$M = 30{\cdot}80$ $\quad ci = -{\cdot}240 \times 5 = -1{\cdot}20$

7নং টেব্‌ল :—শ্রেণীবদ্ধ নম্বরের Mean নির্ণয় (সংক্ষিপ্ত পদ্ধতি)

(খ) Median (Mdn)

(১) যখন নম্বরগুলি শ্রেণীবিহীন অবস্থায় থাকে তখন তাহাদিগকে মানের ক্রমানুসারে সাজাইয়া মধ্যবিন্দু বাহির করিলেই Median পাওয়া যায়। শ্রেণীবিহীন অবস্থায় Mdn বাহির করিতে দুই প্রকার অবস্থার সম্মুখীন হওয়া সম্ভবপর—(ক) যখন N যুগ্ম এবং (খ) যখন N বিযুগ্ম।

যখন N বিযুগ্ম তখন মধ্যবিন্দু বাহির করা সহজ। যেমন কোন অভীক্ষায় 9 জন ছাত্রের ফল মানের ক্রমানুসারে সাজাইলে নিম্নরূপ দাঁড়ায়।

$$4 \quad 4 \quad 6 \quad 7 \quad \overset{*}{8} \quad 9 \quad 10 \quad 11 \quad 13$$

এখানে স্পষ্টই দেখা যাইতেছে যে Mdn 8. স্থান হিসাবে ইহা ঠিক মাঝখানে আছে। তাহার নীচে 4 জন ও উপরে 4 জন আছে।

এখানে যদি প্রথম নম্বর 4 টিকে বাদ দেওয়া হয় তবে 8 জনের ফল হইবে নিম্নরূপ।

$$4 \quad 6 \quad 7 \quad 8 \quad \overset{8{\cdot}5}{\uparrow} \quad 9 \quad 10 \quad 11 \quad 14$$

এখানে Mdn 8 এবং 9 এর মাঝামাঝি কোন বিন্দু। কারণ 8 অথবা 9 Mdn ধরিলে একদিকে বেশী এবং অন্যদিকে কম সংখ্যক ছাত্রের ফল থাকে। তাই 8 এবং 9 এর মাঝামাঝি অর্থাৎ 8·5 ই হইবে Mdn.

শ্রেণীবিহীন অবস্থায় ফলগুলিকে মানের ক্রমানুসারে সাজাইয়া নিম্নলিখিত সূত্রের সাহায্যে Mdn বাহির করা হয়।

$$\text{Median} = \frac{N+1}{2} \text{ তম পদ।} \cdots\cdots\cdots\cdots \qquad (4)$$

প্রথমটিতে N যখন বিযুগ্ম ছিল তখন $\frac{9+1}{2}=5$ তম পদ অর্থাৎ 8ই Median এবং পরেরটিতে N যখন যুগ্ম ছিল তখন $\frac{8+1}{2}=4{\cdot}5$ তম পদ অর্থাৎ 4 তম এবং 5 তমের মাঝামাঝি পদটি অর্থাৎ 8·5ই Median হইল।

(২) যখন যতিবিহীন নম্বরগুলি শ্রেণীবদ্ধ অবস্থায় থাকে তখন Mdn বলিতে সংগা অনুযায়ী গুচ্ছের 50 তম বিন্দুটিকেই নির্দেশ করে। তাই Mdn বাহির করিবার সময় আমরা N/2 কত আগে স্থির করি এবং নীচ বা উপর হইতে গণিয়া যাই। যেখানে 50 তম বিন্দুটি পাই সেই থানেই থামিয়া

যাই। নীচ হইতে গণিয়া N/2 যে ধাপে আছে সেই ধাপের নীচে 8 নং. টেব্‌লের ন্যায় দাগ দেওয়া হয়। নীচ হইতে গণিয়া দেখা গেল যে 30—34 ধাপে Mdn. অবস্থিত। নীচ হইতে ঐ ধাপের পূর্ব ধাপ পর্য্যন্ত f এর যোগফল 20। N/2অর্থাৎ 25 হইতে আর (25—20) = 5 দরকার। এখন ঐ 5টি f পরবর্তী ধাপ অর্থাৎ 30—40এর মধ্যে আছে। কিন্তু উহাতে মোট 10টি f আছে। সুতরাং 5টি f হইতে আর $\left(\frac{5}{10}\times 5\right) = 2{\cdot}5$ নম্বর দরকার। তাহা হইলে Median হইবে ঐ দাগযুক্ত ধাপের নিম্নসীমা অর্থাৎ 29·5 এবং 2·5 এর যোগফল অর্থাৎ 32. মনে রাখা দরকার যে Mdn একটি বিন্দু—উহা নম্বর নহে।

নম্বর	f	
55—59	1	
50—54	2	
45—49	4	
40—44	5	↓
35—39	8	20
30—34	10	
25—29	6	20
20—24	4	↑
15—19	4	
10—14	2	
5— 9	3	
0— 4	1	
	N = 50	

৪নং টেবল : শ্রেণীবদ্ধ নম্বরের Median নির্ণয়

Mdn এর সূত্রটি নীচে প্রদত্ত হইল।

$$\text{Mdn} = l + \left(\frac{\frac{N}{2} - F}{f_m}\right).\, i \qquad \cdots\cdots\cdots\cdots (5)$$

যখন l—যে ধাপে Mdn আছে সেই ধাপের নিম্নসীমা।

$\frac{N}{2}$ মোট সংখ্যার অর্দ্ধেক।

F—যে ধাপে Mdn. আছে সেই ধাপের পূর্ব পর্যন্ত নীচের সকল ধাপের f এর যোগফল।

f_m—যে ধাপে Mdn আছে সেই ধাপের f-এর সংখ্যা।

এবং i—ধাপ দূরত্ব।

উপরের টেবলে $l = 29{\cdot}5,\ \frac{N}{2} = 25,\quad F = 20,\quad f_m = 10,\quad i = 5.$

$$\therefore\quad \text{Mdn} = 29{\cdot}5 + \left(\frac{25-20}{10}\right) \times 5 = 29{\cdot}5 + 2{\cdot}5 = 32{\cdot}00.$$

(গ) Mode—

যখন নম্বরগুলি শ্রেণীবিহীন অবস্থায় থাকে তখন যে নম্বরটি বারবার আসে—তাহাকেই Mode বলে। যে নম্বরটি ঐ দলের ফ্যাশন্—সেই নম্বরটিই ঐ দলের Mode. যেমন 15, 16, 18, 17, 18, 18, 18, 19, 19.—এই দলের মধ্যে 18 বেশীবার আসিয়াছে। তাই ঐ দলের Mode হইতেছে 18. কিন্তু এই প্রকারে যে Mode বাহির হয় তাহা যাথার্থ নহে।

আবার নম্বরগুলি যখন শ্রেণীবদ্ধ ভাবে থাকে তখন যে ধাপে বেশী সংখ্যক f থাকে সেই ধাপের মধ্য বিন্দুই হইতেছে Mode. যেমন 8নং টেবলে 30-34 ধাপে বেশী সংখ্যক f আছে। তাই সেই ধাপের মধ্যবিন্দু অর্থাৎ 32 ঐ সারিগুচ্ছের ফ্যাশন বা Mode. কিন্তু এইভাবে যে Mode পাওয়া যায় তাহা বিশুদ্ধ নহে। উহাকে Crude Mode বলে। যখন সারিগুচ্ছটি Symmetrical, তখন Mode নীচের স্থত্র সাহায্যে বাহির করা হয়।

$$\text{Mode} = 3\ \text{Mdn} - 2\ \text{Mean} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (6)$$

উক্ত 8নং টেবলে $\text{Mdn} = 32{\cdot}00$ এবং $M = 30{\cdot}80$

স্থতরাং $\text{Mode} = 3 \times 32{\cdot}00 - 2 \times 30{\cdot}80$

$$= 96{\cdot}00 - 61{\cdot}60.$$

$$= 34{\cdot}40$$

আপাত দৃষ্টিতে একটি distribution-এর প্রতিভূ হইতেছে Mode. উহা Mean বা Mdn-এর মত সঠিক প্রতিভূ নয়, তাই Mode-এর মান নিখুঁত বিশুদ্ধ হইল কি না হইল তাহাতে কিছু যায় আসে না।

(ঘ) **Mean, Median এবং Mode এর প্রয়োগ :—**

কোন একটি সমস্যার সম্মুখে দাঁড়াইয়া পরিসংখ্যানের ছাত্র কোন্ প্রতিভূ কখন ব্যবহার করিবে তাহা নির্ণয়ে কিছুটা বিব্রত হইয়া পড়ে। সাধারণতঃ Mean কেই প্রতিভূ হিসাবে গ্রহণ করা হয়। কিন্তু এমন অনেক ক্ষেত্র আছে যেখানে Mdn-এর দামই বেশী, Mean সেখানে অচল। সেই অসুবিধা দূর করিবার জন্য নীচে কতকগুলি নিয়ম দেওয়া হইল।

(১) Mean হইতেছে ভরবিন্দু—সকল নম্বরই তাহার ভর-সামঞ্জস্য ঘটায়। সুতরাং যখন নম্বরগুলি কেন্দ্রবিন্দুর দুই পাশে সামঞ্জস্য ভাবে সজ্জিত থাকে,

(২) যখন চরম ও নিরেট সংশুদ্ধ প্রতিভূর প্রয়োজনীয়তা অনুভূত হয়,

এবং (৩) যখন পরিসংখ্যানের অন্যান্য ফল (যেমন SD, r ইত্যাদি) বাহির করিতে হয়,

—তখন Meanই শ্রেষ্ঠ প্রতিভূ হইবে।

আবার, (১) যখন কোন distribution-এর ঠিক শতকরা 50তম বিন্দুটির প্রয়োজন,

(২) যখন সারিতে খুব কম বা খুব বেশী দুই একটি অস্বাভাবিক নম্বর থাকে,

এবং (৩) যখন শ্রেণীবদ্ধ সারির প্রান্তদ্বয়ে কোন নম্বরের উপর বা নীচ বরাবর কিছু f থাকে। কারণ তাহা কত উপর বা কত নীচ না জানা থাকায় সঠিক Mean বাহির করা সম্ভব নয়,

—তখন Mdn. শ্রেষ্ঠ প্রতিভূ হইবে।

আবার, (১) যখন ক্ষিপ্রতার সহিত আনুমানিক প্রতিভূর প্রয়োজন,

(২) যখন আমরা সাধারণে ব্যবহৃত জুতার মাপ, কাপড়ের মূল্য, পরিবারে শিশুর সংখ্যা প্রভৃতির হিসাব করি,

—তখন Mode প্রতিভূ হইবে। কোন কারখানায় নিযুক্ত লোকদের মাহিনার প্রতিভূ হইবে সকলের মাহিনার Modal মান।

Median পরিমাপে অসুবিধা

শ্রেণীগুচ্ছের এক বা একাধিক ধাপে যখন f-এর সংখ্যা শূন্য থাকে তখন Median বাহির করিতে কিছু অসুবিধা ভোগ করিতে হয়। নীচে 8A নং টেবলের সাহায্যে তাহা বুঝান হইতেছে। যেহেতু $N/2 = 10$, নীচ

হইতে Mdn-এর জন্য গণনা করিলে পূর্ব নিয়মানুসারে (5 নং সূত্র) Mdn-এর মান হয়—49·5. কারণ উহা 45-49 ধাপের উচ্চ-সীমা বা 50-54 ধাপের নিম্ন-সীমা। কিন্তু আমরা যদি উপর হইতে গণনা আরম্ভ করি, তবে Median-এর মান দাঁড়ায় 59·5 ; কারণ উহা 60-64 এর নিম্ন-সীমা বা 55-59-এর উচ্চ সীমা।

নম্বর	f	
80—84	4	
75—79	2	
70—74	0	
65—69	0	
60—64	4	55—64
55—59	0	
50—54	0	45—54
45—49	4	
40—44	2	
35—39	2	
30—34	2	
	N = 20	

$$\frac{N}{2} = 10$$

$$\text{Median} = 54{\cdot}5 + \frac{0}{4} \times 5 = 54{\cdot}5$$

8A নং টেব্‌ল ঃ Median নির্ণয়।

এই অসুবিধা হইয়াছে 55-59 এবং 50-54 ধাপদ্বয়ে শূন্য থাকার জন্য। এখন এই দুইটি ধাপের একটিকে উপরের দলে এবং অন্যটিকে নীচের দলে বিভক্ত করিয়া দুইটি নূতন ধাপ-দূরত্ব তৈরী হইল। একটি 45-54 এবং অপরটি 55-64 এবং ইহাদের প্রত্যেকের f-সংখ্যা 4. এইরূপে এখানে চারটি ধাপের পরিবর্তে দুইটি ধাপ হইল। এখন যদি 5নং সূত্রানুযায়ী Mdn. বাহির করা হয়, তাহা হইলে সহজেই তাহার মান 54·5 পাওয়া যাইবে। কারণ 45-54-এর উচ্চ-সীমা অথবা 55-64-ধাপের নিম্ন-সীমা—54·5. হইবে।

Exercises III

1. Calculate the mean, median, and mode for the following frequency distributions. Use the short method in computing the mean.

(a)

Scores	f
52—53	1
50—51	0
48—49	5
46—47	10
44—45	9
42—43	14
40—41	7
38—39	8
36—37	6
34—35	5
32—33	3
Sum	68

(b)

Scores	f
95—99	6
90—94	11
85—89	16
80—84	7
75—79	9
70—74	8
65—69	2
60—64	3
55—59	2
50—54	1
Sum	65

(c)

Scores	f
70—71	2
68—69	2
66—67	3
64—65	4
62—63	6
60—61	7
58—59	5
56—57	4
54—55	2
52—53	3
50—51	1
	N=39

(d)

Scores	f
90—94	2
85—89	2
80—84	4
75—79	8
70—74	6
65—69	11
60—64	9
55—59	7
50—54	5
45—49	0
40—44	2
	N=56

(e)

Scores	f
120—122	2
117—119	2

(f)

Scores	f
100—109	5
90—99	9

(e)

Scores	f
114—116	2
111—113	4
108—110	5
105—107	9
102—104	6
99—101	3
96— 98	4
93— 95	2
90— 92	1
	N=40

(f)

Scores	f
80— 89	14
70— 79	19
60— 69	21
50— 59	30
40— 49	25
30— 39	15
20— 29	10
10— 19	8
0— 9	6
	N=162

(g)

Scores	f
120—139	50
100—119	150
80— 99	500
60— 79	250
40— 59	50
	N=1000

(h)

Scores	f
15	1
14	2
13	3
12	6
11	12
10	15
9	22
8	31
7	18
6	6
5	2
4	2
	N=120

2. Compute the mean and the median for each of the two distributions in 5 (a), Exercises—I, tabulated in 3 and 5 unit intervals. Compare the two means and the two medians, and explain any discrepancy found. [Let the first interval in the first distribution be (61—63); the first interval in the second distribution, 60—64].

3. Compute the arithmetic mean of all distributions in data given below (a to d) inclusive, using the method that seems most feasible. In data c, you will need to make some

assumption about the cases in the two highest intervals. State your assumptions if means are computed for these distributions.

	a	b	c	d
A. Scores	f	f	f	f
66—71	1	0	0	0
60—65	6	0	3	0
54—59	13	1	2	1
48—53	13	1	10	2
42—47	17	3	11	1
36—41	33	3	12	1
30—35	32	2	8	8
24—29	32	9	11	17
18—23	23	16	5	26
12—17	24	36	2	40
6—11	7	43	2	49
0— 5	1	25	0	21

B. Scores	I	II
95—99	1	0
90—94	1	0
85—89	7	1
80—84	13	2
75—79	37	6
70—74	75	23
65—69	189	34
60—64	297	94
55—59	406	144
50—54	425	208
45—49	341	209
40—44	174	205
35—39	81	105
30—34	16	34
25—29	5	15
20—24	0	2
15—19	1	0

C.

Age at Last birthday.	Men	Women
31—35	1	2
26—30	3	6
25	7	6
24	6	7
23	11	7
22	20	6
21	23	16
20	40	13
19	88	48
18	117	67
17	69	57
16	2	6
Sum......	387	241

D.

Scores.	Men	Women
8·0—8·4	1	0
7·5—7·9	5	0
7·0—7·4	2	0
6·5—6·9	7	2
6·0—6·4	6	4
5·5—5·9	11	3
5·0—5·4	10	9
4·5—4·9	16	7
4·0—4·4	18	15
3·5—3·9	19	12
3·0—3·4	17	15
2·5—2·9	17	13
2·0—2·4	14	14
1·5—1·9	13	10
1·0—1·4	8	1
0·5—0·9	1	0
Sums......	165	105

4. The same test is given to the three sections of Grade VI. Results are Section I, M=24, N=32 ; Section II, M=31 ; N=54 ; Section III, M=35, N=16. What is the general mean for the grade ?

5. Compute the median of the following 8 scores.

Scores	f
25—27	1
22—24	1
19—21	2
16—18	0
13—15	2
10—12	0
7—9	2
	N = 8

1. Answers :—

Data.	a	b	c	d	e	f	g	h
Mean	41·7	81·7	60·76	67·36	106·00	55·43	87·5	8·85
Median	41·9	84·7	60·79	66·77	105·83	55·17	87·5	8·55
Mode	42·5	87	60·85	65·59	105·49	54·65	87·5	8·95

2. Class interval = 3 — Mean = 72·92, Median = 71·75
   Class interval = 5 — Mean = 73·00, Median = 72·71

3.

	A				B	
Data	a	b	c	d	I	II
Mean	32·8	13·9	37·2	15·3	54·8	49·3
Median	32·1	11·8	38·0	13·4	54·4	48·8
Mode	38·5	8·5	38·5	8·5	52	47

	C		D	
Data	Men	Women	Men	Women
Mean	19·6	19·7	3·91	3·57
Median	19·1	18·9	3·78	3·43
Mode	18·5	18·5	3·7	3·7

4. 29·43
5. 17·00

চতুর্থ অধ্যায়

বিস্তৃতির পরিমাপ

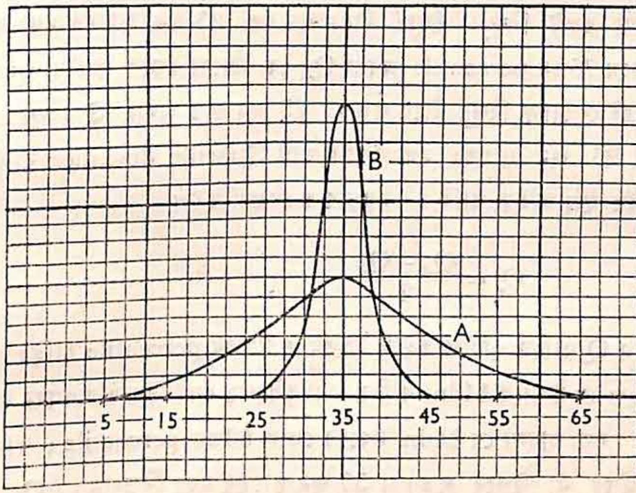
প্রথম অধ্যায়ে কাঁচা নম্বরগুলিকে বিভিন্ন সারি-গুচ্ছে বিভক্ত এবং তাহাদিগকে শ্রেণীবদ্ধ করিয়া কাঁচানম্বরগুলির সারাংশ তৈরী করা হইয়াছে। তৃতীয় অধ্যায়ে সেই শ্রেণীবদ্ধ নম্বরগুলির কেন্দ্রমুখী ঝোঁক লক্ষ্য করিয়া সমস্ত নম্বরগুলির প্রতিভূ Mean, Mdn. এবং Mode বাহির করা হইয়াছে। বর্ত্তমান অধ্যায়ে ঐ প্রতিভূর পাশে নম্বরগুলি কিরূপে সজ্জিত থাকে তাহা আলোচিত হইবে।

শ্রেণীবদ্ধ নম্বরগুলির f কেন্দ্রের দিকে যত বেশী থাকে Mean, Mdn. বা Mode ততই অভ্রান্তভাবে প্রতিভূ হিসাবে গণ্য হইয়া থাকে। কিন্তু কেন্দ্র হইতে দু'পাশে বেশী ছড়ান হইলে, উহারা কিরূপে নিঃসন্দেহে প্রতিভূ হইতে পারে? শেষোক্ত সময়ে সেই প্রতিভূর উপর আমরা আস্থা রাখিতে পারি না। তাই দলের সারাংশ বাহির করিতে কেবলমাত্র প্রতিভূ বাহির করিলেই চলিবে না—অন্যান্য নম্বর প্রতিভূর দুই পাশে কিভাবে বিস্তৃত তাহাও বাহির করা দরকার। কারণ ঐ বিস্তৃতির মানের উপরেই ঐ প্রতিভূর নির্ভরযোগ্যতা নির্ভর করিতেছে।

ধরা হইল, কোন বিদ্যালয়ে একটি বুদ্ধির অভীক্ষা দেওয়া হইল। 50 জন বালক ও 50 জন বালিকা পরীক্ষা দিল। দেখা গেল বালকের ও বালিকাদের গড় নম্বর একই; কিন্তু বালিকারা 25 হইতে 45 এর মধ্যে নম্বর পাইয়াছে (5 নং চিত্রে দ্রষ্টব্য)। বালকেরা 5 হইতে 65 এর মধ্যে নম্বর পাইয়াছে। বালিকাদের নম্বরের (B লেখচিত্র) দূরত্ব 20 (45—25); কিন্তু বালকদের নম্বরের (A লেখচিত্র) দূরত্ব 60 (65—5)। ইহার দ্বারা এই বুঝায় যে বালিকাদের অপেক্ষা বালকদের নম্বরের বিস্তৃতি অনেক বেশী—তাহারা বেশী দূরে ছড়াইয়া আছে—যদিও বালক ও বালিকাদের গড় নম্বর সমান। যদি প্রায় একই মেধাবিশিষ্ট শিক্ষার্থীর উপর অভীক্ষা প্রয়োগ করা হয়, তবে বালিকাদের নম্বরের মত লেখচিত্র এবং বিপরীতপক্ষে বালকদের নম্বরের মত লেখচিত্র হইবে।

নম্বরগুলির বিস্তৃতিকে (Variability বা Dispersion) সাধারণতঃ চার প্রকারে প্রকাশ করা হয়। যথা—(১) Range, (২) Quartile deviation বা Q, (৩) Average deviation অথবা A D এবং (৪) Standard deviation অথবা S D.

(১) Range : 3 নং টেবলে Range এর কথা বলা হইয়াছে এবং নম্বরগুলিকে শ্রেণীবদ্ধ করার সময় ইহাকে কাজে লাগান হইয়াছে। সাধারণতঃ দলের সর্ব্বোচ্চ ও সর্ব্বনিম্ন নম্বরের বিয়োগফলকেই Range বলা হইয়া থাকে। 5 নং চিত্রে বালকদের নম্বরের Range $65-5=60$ এবং



5 নং চিত্র : 50 জন বালক ও 50 জন বালিকার নম্বরের লেখচিত্র।

বালিকাদের Range $45-25=20$. যদি নম্বরগুলি শ্রেণীবদ্ধভাবে বিভিন্ন সারিতে সজ্জিত থাকে তবে প্রথম সারির মধ্যবিন্দু হইতে শেষ সারির মধ্য বিন্দু বিয়োগ করিলে Range পাওয়া যায়। Range দ্বারা বিস্তৃতি বুঝাইতে অসুবিধা আছে। দৈবক্রমে প্রান্তীয় নম্বর দুইটি যদি হঠাৎ ছোট বা বড় হয় তখন Range দ্বারা ঐ distribution-এর প্রকৃত বিস্তৃতি বুঝাইবে না। তাই Range নির্ভরশীল বিস্তৃতি পরিমাপক নয়। কেবলমাত্র সুবিধা এই যে Range দ্বারা শীঘ্র এবং অল্প আয়াসে দলের বিস্তৃতি বাহির করা যায়।

(২) Quartile deviation বা Q :

Range-এর অসুবিধা কি তাহা পূর্বেই আলোচিত হইয়াছে। Range এর এই দুর্বলতা দূর করা সম্ভবপর যদি আমরা প্রান্তীয় নম্বরগুলিকে বাদ দিয়া মাঝের নম্বরগুলির দিকে নজর দিই। আমাদের দৃষ্টি অল্পক্ষেত্রের মধ্যে সীমাবদ্ধ

হইলেও তাহাতে বিস্তৃতির একটি উন্নততর পরিমাপক পাওয়া সম্ভব। কারণ আমরা জানি যে মাঝখানটিকে কেন্দ্র করিয়াই নম্বরগুলি সজ্জিত থাকে।

কোন শ্রেণীগুচ্ছের 75তম এবং 25তম percentile-এর দূরত্বের অর্ধেকের নাম Quartile deviation. এখন 75তম এবং 25তম percentile কি? 25তম percentile অথবা Q_1 কে নম্বরের স্কেলে প্রথম quartile বলে। ইহা শ্রেণীগুচ্ছে একটি বিন্দু। ইহার নীচে শতকরা 25 জন ছাত্রের নম্বর থাকে। সেই রকমে 75তম percentile অথবা Q_3 কে নম্বরের স্কেলে তৃতীয় quartile বলে ; এই শেষোক্ত বিন্দুর নীচে শতকরা 75 জন ছাত্র অর্থাৎ ঐ বিন্দুর উপরে শতকরা 25 জন ছাত্রের নম্বর থাকে তাই Quartile deviation বা Q এর নিম্নলিখিত সূত্র পাইয়া থাকি। নীচে সূত্র দেওয়া হইল।

$$Q = \frac{Q_3 - Q_1}{2} \qquad \text{.........................(7)}$$

এখন Q বাহির করিতে হইলে 75তম ও 25তম percentile বাহির করিতে হইবে এবং যে উপায়ে Mdn. বাহির করা হইয়াছে সেই উপায় অবলম্বন করিতে হইবে। বলা বাহুল্য যে Mdn. নম্বরের স্কেলে 50তম percentile ; কারণ ঐ বিন্দুর নীচে বা উপরে শতকরা 50 জন ছাত্রের নম্বর বিদ্যমান। Q_1 বাহির করিবার সময় 1/4N গণনা করিতে হয় এবং Q_3 বাহির করিবার সময় $\frac{3}{4}$N গণনা করিতে হয় (Mdn. বাহির করিবার সময় $\frac{N}{2}$ গণনা করিতে হয়)। সুতরাং Q_1 এবং Q_3-এর জন্য নূতন দুইটি সূত্র পাওয়া গেল।

$$Q_1 = l + \frac{\frac{N}{4} - \text{cum } f_1}{f_q'} \cdot i \qquad \text{.........................(8)}$$

$$\text{এবং } Q_3 = l + \frac{\frac{3N}{4} - \text{cum } f_1}{f_q} \cdot i \qquad \text{.........................(9)}$$

যেখানে l = যে ধাপে নির্দ্দিষ্ট quartile পড়ে তাহার নিম্ন সীমা।

f_1 = যে ধাপে উক্ত quartile পড়ে তাহার পূর্ব ধাপগুলির f-এর সমষ্টি।

f_q = যে ধাপে ঐ quartile পড়ে সেই ধাপ বরাবর f.

এবং i = ধাপ—দূরত্ব।

নিম্নে একটি উদাহরণ দেওয়া হইল।

নম্বর		f	
110—119		2	
100—109		4	
90—99		6	
80—89		9	
70—79	36	13	
60—69	↑	8	
50—59		3	
40—49		5	12
30—39		3	↑
20—29		2	f_1
10—19		2	
	N = 57		

9নং টেবল : Quartile deviation বা Q নির্ণয়।

(ক) Q_1 নির্ণয় :

9নং টেবলে দেখা যায় যে Q_1 বাহির করিবার সময়

$$l = 49{\cdot}5$$

$$\frac{N}{4} = 14{\cdot}25$$

$$f_1 = 12$$

$$f_q = 3$$

এবং $i = 10$

উপরোক্ত মানগুলি 8নং সূত্রে স্থাপন করিলে

$$Q_1 = 49{\cdot}5 + \frac{14{\cdot}25 - 12}{3} \times 10$$

$$= 49{\cdot}5 + \frac{22{\cdot}5}{3}$$

$$= 49{\cdot}5 + 7{\cdot}5$$

$$= 57{\cdot}00$$

(খ) Q_3 নির্ণয় :

9নং টেবলে দেখা যায় যে Q_3 বাহির করিবার সময়—

$$l = 79{\cdot}5$$
$$\frac{3N}{4} = 42{\cdot}75$$
$$f_1 = 36$$
$$f_q = 9$$
এবং $$i = 10$$

উপরোক্ত মানগুলি 9নং সূত্রে স্থাপন করিলে—

$$Q_3 = 79{\cdot}5 + \frac{42{\cdot}75 - 36}{9} \times 10$$
$$= 79{\cdot}5 + \frac{67{\cdot}5}{9}$$
$$= 79{\cdot}5 + 7{\cdot}5$$
$$= 87{\cdot}00.$$

এখন Q_3 এবং Q_1 এর মান 7নং সূত্রে স্থাপন করিলে

$$Q = \frac{87{\cdot}00 - 57{\cdot}00}{2}$$
$$= 15{\cdot}00.$$

এখানে মনে রাখা দরকার যে সাধারণতঃ Mdn $\pm$ Q এর মধ্যে শ্রেণীগুচ্ছের মাঝের 50% জন ছাত্রের নম্বর থাকে।

(৩) Average deviation (A D)

অথবা Mean deviation (M D) নির্ণয় :

(ক) শ্রেণীবিহীন নম্বরের AD বা MD নির্ণয় :—

কোন দলের Mean হইতে সকল নম্বরের দূরত্বের (deviation) গড়কে AD বা MD বলে। কখনও কখনও Mean এর স্থলে Mdn বা Mode ও লওয়া হইয়া থাকে। AD বা MD বাহির করিবার সময় চিহ্নের প্রতি নজর দেওয়া হয় না। সব deviation বা দূরত্বকে '+' হিসাবে গণ্য করা হয়। নীচে একটি উদাহরণ দেওয়া হইল। 5টি নম্বর 10, 12, 14, 16, এবং 18 এর গড় 14. এখন Mean 14 হইতে প্রতি নম্বরের দূরত্ব যথাক্রমে $10 - 14 = -4$, $12 - 14 = -2$, $14 - 14 = 0$, $16 - 14 = +2$ এবং $18 - 14 = +4$. এখন চিহ্ন বাদ দিয়া সকল দূরত্বের গড় $(4 + 2 + 0 + 2 + 4)$

$\div 5 = 2{\cdot}4$. $\therefore$ ঐ পাঁচটি নম্বরের শ্রেণীর AD বা MD = 2·4. সূত্র নীচে দেওয়া হইল। যখন নম্বরগুলি শ্রেণীবিহীন তখন

$$\text{AD বা MD} = \frac{\Sigma|x|}{N} \quad \cdots\cdots\cdots\cdots (10)$$

উপরের সূত্রে x এর অর্থ হইতেছে কোন নম্বর (X) এবং Mean (M) এর বিয়োগ ফল (x = X − M). x এর দুই পাশে দুইটি সোজা প্রাচীর দেওয়া হইয়াছে। এর অর্থ এই যে deviation নির্ণয়ে এখানে চিহ্নের দিকে লক্ষ্য দেওয়া হইবে না। সব x কেই ধনাত্মক হিসাবে লওয়া হয়।

(খ) শ্রেণীবদ্ধ নম্বরের AD বা MD নির্ণয় :—

সাধারণতঃ AD বর্ত্তমান সময়ে পরিসংখ্যানে বড় একটা চলে না। পূর্বে বিভিন্ন পুস্তকে AD বা MD এর নজীর মিলে। যাহা হউক শ্রেণীবদ্ধ নম্বরের AD বাহির করার নিয়ম নিম্নে প্রদত্ত হইল।

নম্বর	f	মধ্যবিন্দু		x	fx	fx²
55—59	1	57	+	26·20	26·20	686·44
50—54	2	52	+	21·20	42·40	898·88
45—49	4	47	+	16·20	64·80	1049·76
40—44	5	42	+	11·20	56·00	627·20
35—39	8	37	+	6·20	49·60	307·52
30—34	10	32	+	1·20	12·00	14·40
25—29	6	27	—	3·80	—22·80	86·64
20—24	4	22	—	8·80	—35·20	309·76
15—19	4	17	—	13·80	—55·20	761·76
10—14	2	12	—	18·80	—37·60	706·88
5— 9	3	7	—	23·80	—71·40	1699·32
0— 4	1	2	—	28·80	—28·80	829·44
	N = 50				Σ|x| = 502·00	Σfx² = 7978·00

10 নং টেবল : শ্রেণীবদ্ধ নম্বরের AD নির্ণয়

$$M = 30{\cdot}80$$

$$\text{AD বা MD} = \frac{\Sigma|fx|}{N} \quad \cdots\cdots\cdots\cdots (11)$$

$$= \frac{502}{50}$$

$$= 10{\cdot}04.$$

প্রথমে প্রতি সারির f বরাবর মধ্যবিন্দুগুলিকে বাহির করিতে হয়। 10 নম্বর টেবলে Mean নম্বর হইতেছে 30·80. এখন x স্তম্ভ ভর্ত্তি করিতে হইবে। প্রতি মধ্য বিন্দু Mean হইতে কত দূরে তাহা নির্ণয় করিয়া x (X—M) স্তম্ভে রাখিতে হইবে। চিহ্নের দিকে লক্ষ্য না রাখিয়া সব deviation এর সাম্নে '+' চিহ্ন দেওয়াও যাইতে পারে। পরে fকে f বরাবর x দ্বারা গুণ করিয়া fx স্তম্ভ পূর্ণ করিতে হইবে। fx স্তম্ভ যোগ করিয়া N দিয়া ভাগ করিলেই AD বা MD বাহির হইবে।

পূর্বে বলা হইয়াছে যে Q মাঝের 50% জন ছাত্রের খবর দেয়। আর M ± AD সীমানার মধ্যে 57·5% ছাত্রের নম্বর থাকে। সুতরাং Q হইতে AD দ্বারা মধ্যস্থলের অপেক্ষাকৃত বেশী দূরত্বের খবর জানা যায়।

(৪) Standard deviation বা SD.

পূর্বেই বলা হইয়াছে MD বা AD এর প্রচলন কেবলমাত্র পুস্তকেই সীমাবদ্ধ থাকে। ইহার পরিবর্ত্তে SD প্রচলিত হইয়াছে। তাহা ছাড়া SD পরিসংখ্যানে পরবর্ত্তী বহু সূত্র নির্ণয়ে সাহায্য করিয়াছে। MD বাহির করিবার সময় কেন '—' চিহ্নের বদলে সব deviation এর পূর্বে '+' চিহ্ন দেওয়া হয়, তাহার কোন যুক্তি নাই। SDতে deviation গুলির বর্গ নির্ণয় করিয়া '—' চিহ্নের হাত হইতে রক্ষা পাওয়া গিয়াছে। $(-2)^2$ এর মান এবং $(+2)^2$ এর মান উভয় ক্ষেত্রেই + 4.

Mean হইতে নম্বরগুলির দূরত্বের বর্গ নির্ণয় করিয়া গড় বাহির করিলে পরিসংখ্যানে Variance পাওয়া যায়। Variance এর বর্গমূল বাহির করিলে Standard deviation বা SD পাওয়া যায়।

তাহা হইলে কোন distributionএ Mean হইতে সকল নম্বরের দূরত্বের বর্গের গড়ের বর্গমূলই হইল সেই distribution এর SD. অবশ্য যদি শ্রেণীবদ্ধ থাকে তাহা হইলে ঐ বর্গমূলকে সারি-দূরত্ব (i) দ্বারা গুণ করিলে তবে SD পাওয়া যাইবে। SDকে গ্রীক শব্দ σ দ্বারা সূচিত করা হয়। নীচে σ বাহির করার নিয়ম প্রদত্ত হইল।

(ক) যখন নম্বরগুলি শ্রেণাবদ্ধ নয় :

আমরা AD বাহির করিবার সময় 5 জন ছাত্রের নম্বর লইয়াছিলাম। SD বাহির করিবার সময় সেই 5টি ছাত্রের নম্বর লইতেছি। নম্বরগুলি যথাক্রমে 10, 12, 14, 16, এবং 18. ইহাদের Mean 14. প্রতিটি deviation এর

বর্গ যথাক্রমে 16, 4, 0, 4 এবং 16. উহাদের গড় $(16+4+0+4+16)\div 5 = 8$. এখন σ হইতেছে $\sqrt{8} = 2{\cdot}83$. কারণ

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

(খ) যখন নম্বরগুলি শ্রেণীবদ্ধ অবস্থায় থাকে তখন SD বাহির করিবার নিয়ম 10 নং টেবলে দেখান হইয়াছে। AD বাহির করিবার সময় fx স্তম্ভ কাজে লাগিয়াছে। এখন σ বাহির করিবার সময় fx^2 স্তম্ভটি বাহির করিতে হইবে। xকে x বরাবর fx স্তম্ভের fx দ্বারা গুণ করিলে fx^2 পাওয়া যায়। fx^2 স্তম্ভে x এর বর্গ হওয়ার দরুণ '—' চিহ্ন লোপ পায়।

এখন নীচের সূত্রের সাহায্যে σ বাহির করা হয়।

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

(যখন নম্বরগুলি শ্রেণীবদ্ধ)

10নং টেবলে fx^2 স্তম্ভের যোগফল 7978·00 এবং N=50.

$$\therefore \quad \sigma = \sqrt{\frac{7978{\cdot}00}{50}}$$

$$= \sqrt{159{\cdot}56}$$

$$= 12{\cdot}63.$$

(গ) নম্বরগুলি যখন শ্রেণীবদ্ধ থাকে তখন উপরোক্ত দীর্ঘ এবং জটিল পদ্ধতিতে অগ্রসর না হইয়া নিম্নলিখিত সূত্রের সাহায্যে ও নূতন পদ্ধতিতে σ বাহির করা যায়। তাহাতে শক্তি ও সময়ের সদ্ব্যবহার হয়।

$$\sigma = i\sqrt{\frac{\Sigma fx^{1^2}}{N} - c^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

যখন x^1 = X – Assumed Mean.

c=Assumed Mean লওয়ার জন্য সংশোধন।

i=ধাপ–দূরত্ব।

এবং N = মোট পরীক্ষার্থী

নম্বর	f	x^1	fx^1	${fx^1}^2$
55—59	1	5	5	25
50—54	2	4	8	32
45—49	4	3	12	36
40—44	5	2	10	20
35—39	8	1	8	8
30—34	10	0	0	0
25—29	6	—1	— 6	6
20—24	4	—2	— 8	16
15—19	4	—3	—12	36
10—14	2	—4	— 8	32
5— 9	3	—5	—15	75
0— 4	1	—6	— 6	36
	N = 50		Σfx^1 = —12	$\Sigma {fx^1}^2$ = 322

11নং টেবল : শ্রেণীবদ্ধ নম্বরের σ নির্ণয় (সংক্ষিপ্ত পদ্ধতি)

সংক্ষিপ্ত পদ্ধতিতে σ বাহির করিতে Mean বাহির করার মত শ্রেণীর মাঝখানে বা সর্বাপেক্ষা বেশী f এর ধাপে আনুমানিক Mean লইতে হয়। Mean বাহির করিবার পদ্ধতির অনুরূপ x^1 এবং fx^1 স্তম্ভ দুইটি ভর্ত্তি করিতে হয়। এখানে অতিরিক্ত হইতেছে ${fx^1}^2$ স্তম্ভটি। x^1 কে fx^1 দ্বারা বরাবর গুণ করিয়া ${fx^1}^2$ স্তম্ভটি পাওয়া যায়। এখানে $\Sigma {fx^1}^2 = 322$. এখন 14নং সূত্রানুযায়ী σ এর মান বাহির করা হয়।

$$\sigma = i\sqrt{\frac{\Sigma {fx^1}^2}{N} - c^2}$$

$$= 5\sqrt{\frac{322}{50} - .0576} \qquad [\because c = -\cdot 24]$$

$$= 12{\cdot}63.$$

এখানে লক্ষ্য করা উচিত যে দীর্ঘ পদ্ধতি এবং সংক্ষিপ্ত পদ্ধতিতে σ এর মান দুই ক্ষেত্রেই 12·63 হইবে।

(ঘ). কাঁচা নম্বর হইতে σ বাহির করিবার নিয়ম :

শ্রেণীবদ্ধ গুচ্ছের ন্যায় শ্রেণীবিহীন কাঁচা নম্বর হইতে সংক্ষিপ্ত পদ্ধতিতে σ

বাহির করা যায়। নম্বরগুলিকে ক্রমানুসারে সাজাইবার দরকার নাই। একটি স্তম্ভে যে কোন স্থানে তাহাদিগকে রাখা হইল। মনে করা গেল যে Assumed Mean = 0 তাহা হইলে deviation (x^1) এবং নম্বর (X) একই রহিল। এখন $c = M - 0 = M$. আবার $c = \frac{\Sigma x^1}{N} = \frac{\Sigma X}{N} = M$. তাহা হইলে নূতন সূত্রটি এইরূপ দাঁড়ায়—

$$\sigma = \sqrt{\frac{\Sigma {x^1}^2}{N} - c^2}$$

$$= \sqrt{\frac{\Sigma X^2}{N} - M^2} \qquad \cdots\cdots\cdots\cdots(15)$$

$$= \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$$

$$= \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N}} \qquad \cdots\cdots\cdots\cdots(16)$$

নীচে 12নং টেবলে একটি উদাহরণ দেওয়া গেল।

নম্বর (x)	x^1 (বা X)	$(x^1)^2$ বা (X^2)
13	13	169
12	12	144
10	10	100
16	16	256
15	15	225
11	11	121
14	14	196
18	18	324
109	109	1535

12নং টেবল : শ্রেণীবিহীন কাঁচা নম্বরের σ নির্ণয় (সংক্ষিপ্ত পদ্ধতি)

12নং টেবলে AM = 0.

$$M = \frac{109}{8} = 13{\cdot}63 \quad (N=8)$$

$$c = 13{\cdot}63$$

$$c^2 = 185{\cdot}78$$

$$\therefore \quad \sigma = \sqrt{\frac{1535}{8} - 185{\cdot}78}$$

$$= \sqrt{191{\cdot}88 - 185{\cdot}78}$$

$$\doteq 2{\cdot}47.$$

কেবলমাত্র একটি calculating machine এবং বর্গের তালিকা লইয়া উক্ত অঙ্ক অল্পায়াসে করা যায়।

COEFFICIENT OF VARIATION—V

শ্রেণীর নম্বরের যে একক, σ বা Q-কে সেই এককেই প্রকাশ করা হয়। যেমন দৈর্ঘ্যের σ তাহার M এর মত দৈর্ঘ্যের এককে পরিণত হয়—আবার ওজনের σ তাহার M এর মত ওজন (মন, সের, পাউণ্ড ইত্যাদি) এককে পরিণত হয়। যদি দুই রকম এককের M একই বা কাছাকাছি হয় তবে তাহাদের σ দুইটি দেখিয়া দল দুইটির বিস্তৃতির মধ্যে তুলনা করা সম্ভব। কোন একটি (A) দলের $M=31$ ধরা যাক্ এবং তাহার $\sigma=13$. অন্য একটি (B) দলের $M=29$ এবং তাহার $\sigma=3$. তাহা হইলে স্পষ্টই বুঝা যাইতেছে যে B দল অপেক্ষা A দলের ফলের বিস্তৃতি বেশী—প্রায় চারি গুণ।

আবার অনেক সময় এমনও দেখা যায় যে দুই দলের Mean বিভিন্ন অথবা প্রথমটির M এবং σ এক এককে প্রকাশিত। ভিন্ন একটির M এবং σ অন্য এককে প্রকাশিত। তখন কিরূপে দুই দলের বিস্তৃতির তুলনা সম্ভবপর? সেই সময় দুই দলের বিস্তৃতিকে এক Standard এককে প্রকাশ করিয়া তুলনা করা হয়। বিস্তৃতির সহগ বা coefficient of Variation বা V দ্বারা সেই তুলনা সম্ভবপর। V একক নিরপেক্ষ।

$$V = \frac{\sigma}{M} \times 100 \quad \cdots\cdots\cdots\cdots\cdots\cdots(17)$$

প্রাকৃতিক পরিমাপে বিশেষতঃ পরিমান, ওজন, সময় প্রভৃতির তুলনার সময় V ব্যবহার করা যাইতে পারে। Ratio Scale এ দুই প্রকার অবস্থা থাকিতে পারে—(১) যখন দুই দলের M এবং σ এর একক বিভিন্ন এবং (২) দুই দলের M এবং σ এর একক এক ; কিন্তু Mean বিভিন্ন।

(১) **যখন একক বিভিন্ন :** 9 বৎসর বয়স্ক কোন এক দল ছাত্রের Mean উচ্চতা = 48″ ইঞ্চি এবং উচ্চতার $\sigma=3''$ ইঞ্চি। তাহাদের Mean ওজন= 54 পাউণ্ড এবং ওজনের $\sigma=7$ পাউণ্ড।

$$\therefore \quad V_{ht}=\frac{3}{48} \times 100 = 6{\cdot}25$$

$$V_{wt}=\frac{7}{54} \times 100 = 12{\cdot}96$$

এর দ্বারা এই সিদ্ধান্তে আসা যায় যে বালকেরা উচ্চতার তুলনায় ওজনে দ্বিগুণ Variable.

(12·96 ÷ 6·25 = 2·07)

(২) **যখন একক একই, কিন্তু MEAN বিভিন্ন ঃ** নীচে পুরুষ ও বালকের ওজন দেওয়া হইল।

দল	M	σ	V
পুরুষ	167	17	10·19
বালক	45	5	11·11

দুই দলের σ এর দিকে তাকাইলে মনে হয় যেন পুরুষ বালক অপেক্ষা প্রায় তিনগুণ Variable, কিন্তু তাহারা যখন Mean এর অনুপাতে আসিল তখন প্রায় সমান Variable হইয়াছে। প্রথমটিতে V হইয়াছে 10·19 এবং দ্বিতীয়টিতে হইয়াছে 11·11.

V ব্যবহারে সতর্কতা

মানসিক ও শিক্ষাগত মানের তুলনা V দ্বারা সূচিত করিতে বিভিন্ন দিক হইতে আপত্তি উঠিয়াছে। মানসিক অভীক্ষাগুলি অনুপাত স্কেল নয়—যেখানে absolute 0 অজ্ঞাত। তাই V দ্বারা তুলনা করিলে ভুল হইবে। সকলের জ্ঞাত 10টি প্রশ্ন অভীক্ষায় দিলে M হয়ত 10 বৃদ্ধি পাইবে (প্রতি প্রশ্নের মান যদি 1 একক হয়) ; কিন্তু σ পূর্ববৎই থাকিবে। তাই সাধারণতঃ একই অভীক্ষা বিভিন্ন পরিবেশে একই ছাত্রদলের উপর প্রয়োগ করিলে সেইখানে দুইটি দলের তুলনা V দ্বারা সম্ভবপর। অথবা দুইটি দলের Mean ফল যখন প্রায় সমান তখন একই অভীক্ষা দুই দলের উপর প্রয়োগ করিয়া V দ্বারা তুলনা করা চলে। একই ছাত্রের উপর দুই রকম অভীক্ষার (মনে করি পঠন ও গণিত) তুলনা V দ্বারা করিলে সেই তুলনা ধ্রুবক হইবে না ; কেবলমাত্র সেই অভীক্ষা দুইটির পক্ষে প্রযোজ্য হইবে। অন্য পঠন ও গণিতের অভীক্ষার তুলনার পক্ষে কার্যকরী নয়।

বিস্তৃতি পরিমাপের ব্যবহার

(১) Range ব্যবহার করিতে হইবে,

(ক) যখন data এত ক্ষুদ্র বা এত বেশী ছড়ান যে অন্য কোন Measure of variability-র প্রয়োজন হয় না।

(খ) যখন প্রান্তীয় নম্বর অথবা মোট বিস্তৃতির জ্ঞান দরকার হয়।

(২) Q ব্যবহার করিতে হইবে,

(ক) যখন Mdn ই দলের প্রতিভূ

(খ) যখন distributionটি বেশ ছড়ান বা প্রান্তে এমন সব নম্বর থাকে যাহারা σ কে অযথা প্রভাবান্বিত করে।

(গ) যখন মাঝের 50% এর খবর আমরা চাই।

(৩) AD ব্যবহার করিতে হইবে,

(ক) যখন সকল নম্বরের সব বিস্তৃতির পরিমাপ করি।

(খ) যখন প্রান্তীয় বিস্তৃতি σ কে অযথা প্রভাবান্বিত করে।

(৪) SD বা σ ব্যবহার করিতে হইবে,

(ক) যখন পরিসংখ্যানে সর্বাপেক্ষা বেশী সত্যতা নিরূপণের দরকার হয়।

(খ) যখন প্রান্তীয় বিস্তৃতিগুলি Variabilityর উপর আনুপাতিকভাবে বেশী প্রভাব বিস্তার করে।

(গ) যখন অনুসঙ্গসহগ এবং পরিসংখ্যানের অন্যান্য তথ্য বাহির করিবার দরকার হয়।

Exercises—IV

1. Calculate the Q and σ for each of the four frequency distributions given in 1, Exercises—III.

2. For the following list of test scores,

52, 50, 56, 68, 65, 62, 57, 70.

(a) Find the M and σ by any method you like.

(b) Add 6 to each score and recalculate M and σ.

(c) Subtract 50 from each score, and calculate M and σ.

(d) Multiply each score by 5 and compute M and σ.

3. Compute the coefficient of Variation for each of the following distributions.

Test	Tapping Rate		Hand grip		Steadiness	
	Men	Women	Men	Women	Men	Women
Mean	210·4	184·0	42·1	23·9	5·64	5·13
Standard deviation	20·0	19·3	6·4	4·8	1·6	1·9
N	101	161	108	172	105	165

4. Calculate coefficients of Variation for the following traits :—

Trait	Unit of Measurement	Group	M	σ
Length of Head	mms.	802 males	190·52	5·90
Body weight	Pounds	868, 445 males	141·54	17·82
Tapping Speed	M of 5 trails 30″ each	68 adults, male & fem.	196·91	26·83
Memory Span	No. repeated correctly	263 males	6·60	1·13
General Intelligence	Points scored	1101 adults	153·3	23·6

Answers :—

1. (a) Q=3·5, σ=4·58 (b) Q=7·7, σ=10·86 (c) Q=3·37, σ=4·99 (d) Q=8·12, σ=11·33
(e) Q=4·50, σ=7·23 (f) Q=16·41, σ=24·13

2. (a) M=60, σ=6·91 (b) M=66, σ=6·91 (c) M=10, σ=6·91 (d) M=300, σ=34·55

3. 9·51 ; 10·5 ; 15·2 ; 20·1 ; 28·4 ; 37·0.

4. V's in order are 3·10 ; 12·59 ; 13·63 ; 17·12 ; 15·39. Ranked for relative variability from most to least. Memory span, General Intelligence ; Tapping Speed , Weight , Head Length. Last two traits have true zeroes.

পঞ্চম অধ্যায়

Cumulative distribution, Graphic Method এবং Percentiles.

দ্বিতীয় অধ্যায়ে Polygon curve এবং Histogram সম্বন্ধে আলোচনা করা হইয়াছে। এই অধ্যায়ে cumulative frequency curve এবং cumulative percentage curve বা ogive সম্বন্ধে আলোচনা করা হইবে। ইহা ছাড়া অন্যান্য দুই একটি প্রয়োজনীয় curve এরও আলোচনা থাকিবে।

A. CUMULATIVE FREQUENCY CURVE

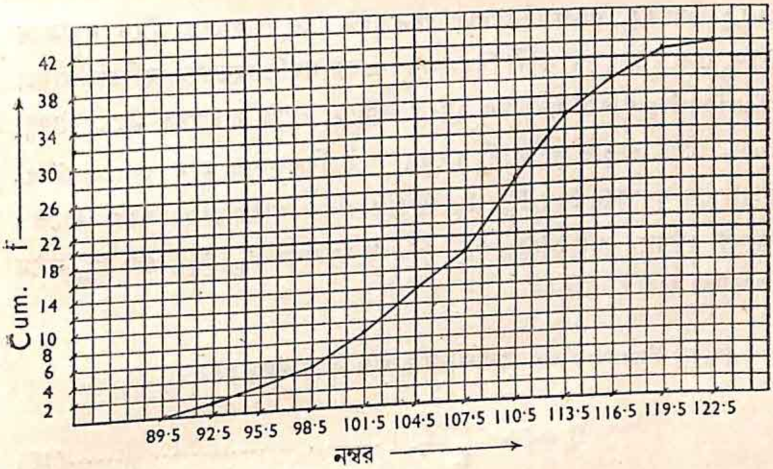
Frequency distributionকে cumulative frequency curve দ্বারা অন্যরূপে প্রকাশ করা যায়। ঐ লেখচিত্র অঙ্কনের পূর্বে 13নং টেবলে যেরূপভাবে f স্তম্ভের যোগফল দেখান হইয়াছে তেমনি করিয়া cum. f স্তম্ভটি করিতে হইবে। 6নং টেবল হইতে উক্ত ফলটি লওয়া হইয়াছে।

নম্বর	f	cum f.
120—122	1	42
117—119	3	41
114—116	4	38
111—113	7	34
108—110	9	27
105—107	4	18
102—104	5	14
99—101	4	9
96— 98	2	5
93— 95	2	3
90— 92	1	1
87— 89	0	0
	N=42	

13নং টেবল : শ্রেণীবদ্ধ distribution এর cumulative f নির্ণয়।

Cum. f স্তম্ভটি পূর্ণ করিবার সময় নীচ হইতে f গুলি ক্রমান্বয়ে যোগ করিয়া সেই ধাপ বরাবর cum f. বসাইতে হয়। উক্ত টেবলে যেমন 90—92

ধাপের 92·5 পর্যন্ত একটি f আছে; তাই cum f. স্তম্ভে 1 লেখা হইল। আবার 90—92 ধাপ এবং 93—95 এর মধ্যে অর্থাৎ 89·5 (90 এর নিম্ন সীমা) হইতে 95·5 (95 এর উচ্চ সীমা) পর্যন্ত 1+2=3টি f আছে। তাই 93—95 ধাপ বরাবর cum f. স্তম্ভে 3 লেখা হইয়াছে। এইরূপে 96—98 ধাপ বরাবর cum f. স্তম্ভে 5 লেখা হইয়াছে। প্রতিবার নীচ হইতে কোন ধাপ পর্যন্ত f এর যোগ না করিয়া সহজে cum f. বাহির করা যায়। যেমন 90—92 ধাপ বরাবর 1; পরেরটিতে 1+2=3; পরেরটিতে 3+2=5; পরেরগুলি যথাক্রমে 5+4=9; 9+5=14 ইত্যাদি। বলাবাহুল্য মোট f সংখ্যা এই টেব্‌লে 42 থাকার দরুণ সর্বোপরি cum f. হইবে 42. লেখচিত্র অঙ্কনের সুবিধার জন্য 90—92 ধাপের নীচে 87—89 ধাপটি এবং ঐ ধাপ বরাবর f এবং cum f. স্তম্ভে 0 লইতে হয়।



6নং চিত্র : cumulative frequency curve.

Cumulative frequency লেখচিত্র অঙ্কনের সময় frequency polygon বা Histogram এর ন্যায় 75% নীতি অনুসরণ করিতে হয়। Frequency polygonএ f মধ্য বিন্দুর উপর এবং Histogramএ f গোটা ধাপের উপর অঙ্কিত হয়। কিন্তু cumulative frequency curveএ cum f. বরাবর ধাপের উচ্চ সীমার উপরই cum f. বিন্দু বসাইতে হয়। কারণ cum f. কোন ধাপ বিশেষের f নয়। উহা ঐ ধাপ পর্যন্ত f এর সমষ্টি

হওয়ায় ঐ ধাপের উচ্চ সীমার উপর cum f. টির বিন্দু বসাইতে হয়, তাই 89·5 নম্বরের উপর 0, y—একক, 92·5 নম্বরের উপর 1, y—একক, 95·5 নম্বরের উপর 3 y—একক প্রভৃতি এবং শেষ নম্বর 122·5এর উপর 42 y—এককের উপর বিন্দু বসিবে। কোন distribution এর সর্ব নিম্ন এবং সর্বোচ্চ ধাপে যদি 'এত' এর নীচে বা 'এত' এর উপরে থাকে তবে ঐ প্রান্ত বিন্দুদ্বয় যুক্ত না করাই উচিত।

PERCENTILES এবং PERCENTILE RANKS

(ক) বিভিন্ন Percentile নির্ণয় ঃ

আমরা পূর্বেই জানিয়াছি যে Mdn, distribution এর এমন একটি বিন্দু যাহার নীচে শতকরা 50 জন পরীক্ষার্থী বিদ্যমান। আবার ইহাও জানিয়াছি যে Q_1 এবং Q_3, distributionএ এমন দুইটি বিন্দু যাহাদের নীচে যথাক্রমে 25% এবং 75% পরীক্ষার্থী বিদ্যমান। একই পদ্ধতি অবলম্বন করিয়া আমরা distributionএ এমন সব বিন্দু বাহির করিতে পারি যাহাদের নীচে 18%, 37%, 89% প্রভৃতি পরীক্ষার্থী থাকে। এই বিন্দুগুলির নাম percentiles. ইহাদিগকে যথাক্রমে P_{18}, P_{37}, P_{89} প্রভৃতি বলিয়া আখ্যা দেওয়া হইয়া থাকে। আমরা বর্ত্তমানে তাই Mdn., Q_1, Q_3কে যথাক্রমে P_{50}, P_{25}, এবং P_{75} নামে অভিহিত করিতে পারি।

বিভিন্ন Percentile বাহির করিবার স্থত্র নীচে প্রদত্ত হইল।

$$P_p = l + \frac{P_n - F}{f_p} . i \quad \cdots\cdots\cdots\cdots\cdots\cdots (18)$$

(নীচ হইতে উপরে গণনা করিয়া percentile বাহির করার স্থত্র)

যখন p=distribution এর যে শতকরা f দরকার।

$l = P_p$ যে ধাপে আছে তাহার নিম্ন সীমা।

$P_n = P_p$তে পৌছিতে গেলে N এর যে অংশ প্রয়োজন।

$F = l$ এর নীচে অবস্থিত f এর যোগফল।

f_p = যে ধাপে P_p অবস্থিত সেই ধাপের f,

এবং i = ধাপ-দূরত্ব।

নীচে 14নং টেবলের সাহায্যে বিভিন্ন percentile নির্ণয় পদ্ধতি দেখান হইয়াছে।

নম্বর	f	cum. f	Percentiles
90—94	1	120	$P_{100}=94{\cdot}50$
85—89	3	119	$P_{90}=79{\cdot}08$
80—84	7	116	$P_{80}=74{\cdot}28$
75—79	12	109	$P_{70}=71{\cdot}55$
70—74	22	97	$P_{60}=69{\cdot}01$
65—69	31	75	$P_{50}=67{\cdot}09$
60—64	18	44	$P_{40}=65{\cdot}15$
55—59	15	26	$P_{30}=62{\cdot}28$
50—54	7	11	$P_{20}=59{\cdot}83$
45—49	2	4	$P_{10}=54{\cdot}83$
40—44	2	2	$P_0=39{\cdot}50$
	N=120		

14নং টেবল : শ্রেণীবদ্ধ সারির percentile নির্ণয়।

120 এর $10\%=12$ $\quad P_{10}=54{\cdot}5+\frac{12-11}{15}{\cdot}5=54{\cdot}83$

120 এর $20\%=24$ $\quad P_{20}=54{\cdot}5+\frac{24-11}{15}{\cdot}5=59{\cdot}83$

120 এর $30\%=36$ $\quad P_{30}=59{\cdot}5+\frac{36-26}{18}{\cdot}5=62{\cdot}28$

120 এর $40\%=48$ $\quad P_{40}=64{\cdot}5+\frac{48-44}{31}{\cdot}5=65{\cdot}15$

120 এর $50\%=60$ $\quad P_{50}=64{\cdot}5+\frac{60-44}{31}{\cdot}5=67{\cdot}00$(Mdn)

120 এর $60\%=72$ $\quad P_{60}=64{\cdot}5+\frac{72-44}{31}{\cdot}5=69{\cdot}01$

120 এর $70\%=84$ $\quad P_{70}=69{\cdot}5+\frac{84-75}{22}{\cdot}5=71{\cdot}55$

120 এর $80\%=96$ $\quad P_{80}=69{\cdot}5+\frac{96-75}{22}{\cdot}5=74{\cdot}28$

120 এর $90\%=108$ $\quad P_{90}=74{\cdot}5+\frac{108-97}{12}{\cdot}5=79{\cdot}08$

120 এর $100\%=120$

(খ) Frequency distribution এ percentile rank নির্ণয় :

এতক্ষণ শ্রেণীবদ্ধ সারির বিভিন্ন percentile (যথা P_{18}, P_{23}, P_{97} ইত্যাদি বাহির করা হইয়াছে। Percentile যতিবিহীন সারির এমন একটি বিন্দু যাহার নীচে নির্দিষ্ট সংখ্যক শতকরা N বিদ্যমান থাকে। এখন শতের মধ্যে কোন ছাত্রের সঠিক স্থান নির্ণয় করা হইবে। শতের মধ্যে এই স্থানের নাম percentile rank বা PR. তাহা হইলে percentile এবং PR এ তফাৎ কি ? প্রথমতঃ percentile একটি নম্বর ; কিন্তু PR শতের মধ্যে একটি বিন্দু। দ্বিতীয়তঃ percentile বাহির করার সময় N এর শতকরা অংশ লইয়া নম্বর বাহির করা হয় ; কিন্তু PR বাহির করার সময় নম্বর দেওয়া থাকে— শতে তাহার অবস্থান বাহির করিতে হয়। যেমন P_{60} একটি নম্বর ; সে নম্বর 14নং টেবলে 69·01 ; কিন্তু উক্ত টেবলে 79 এর PR প্রায় 90.

14নং টেবলের একটি উদাহরণ লওয়া যাক। 79 নম্বরের PR বাহির করিতে হইবে। 79 নম্বর 75-79 ধাপে আছে। ঐ ধাপের পূর্বে মোট f এর সংখ্যা হইতেছে 97· 75-79 এর নিম্নসীমা 74·5. নম্বর 79 পাইতে হইলে $79-74{\cdot}5=4{\cdot}5$ নম্বর দরকার। এখন অতিরিক্ত 4·5 নম্বরে কত f আছে তাহা বাহির করিতে হইবে। ঐ ধাপে মোট 12টি f আছে। এখন 4·5 নম্বরে ঐ ধাপে $\frac{4{\cdot}5}{5} \times 12$টি $= {\cdot}9 \times 12 = 10{\cdot}8$টি f বিদ্যমান। ∴ 120টি f এর মধ্যে 79 নম্বরের নীচে $(97+10{\cdot}8=)$ 107·8টি f বিদ্যমান। ∴ 79 নম্বরের নীচে 89·8% ছাত্র অবস্থিত। তাই 79 এর PR 90.

(গ) মানের ক্রমানুসারে সজ্জিত নম্বরের PR নির্ণয় :

কখনও কখনও বিষয় বস্তু বা ব্যক্তির বৈশিষ্ট্য বা কোন গুণকে প্রত্যক্ষভাবে মাপা যায় না। তাহাদের স্থানের মান অনুযায়ী 1, 2, 3 ইত্যাদি পদমর্যাদায় সাজান হয়। ধরা যাক 10 জন ব্যক্তির মধ্যে সততা নির্ণয়ে সর্বশ্রেষ্ঠ ব্যক্তিকে 1 এবং সর্বনিম্ন ব্যক্তিকে 10 এইভাবে 10 জনকে যথাক্রমে 1, 2, 3,……10 পদমর্যাদা দেওয়া হইল। এইরূপ মানকে নিম্নলিখিত সূত্রের সাহায্যে percentile rank-এ পরিবর্তিত করা যায়।

$$PR = 100 - \frac{(100\,R - 50)}{N} \quad \ldots\ldots\ldots\ldots\ldots \ (19)$$

সেই জন্য উপরোক্ত যে ব্যক্তির পদমর্যাদা '1' তাহার PR হইবে $100-\frac{100\times 1-50}{10}=95$; যাহার পদমর্যাদা 8 তাহার PR হইবে $100-\frac{100\times 8-50}{10}=25$; এবং সর্বনিম্নের PR হইবে $100-\frac{100\times 10-50}{10}=5$.

B. Cumulative percentage curve বা ogive

Cumulative percentage curve বা ogive, cumulative frequency graph এরই অনুরূপ। কেবল মাত্র frequency গুলিকে শতকরায় পরিবর্ত্তিত করা হইয়াছে ; 15নং টেবলে frequency গুলিকে শতকরায় লইয়া যাওয়া হইয়াছে। ঐ টেবলে 3নং এবং 4নং স্তম্ভে যথাক্রমে cumulative frequency এবং cumulative percentage frequency আছে। এখন 4নং স্তম্ভটি কিরূপে আসিল ? Cumulative frequencyকে N দ্বারা ভাগ করিয়া শতকরায় লইয়া গেলে 3নং স্তম্ভটি 4নং স্তম্ভে পরিণত হইবে। সহজে এবং স্বল্প পরিশ্রমে 4নং স্তম্ভটি 1/N বা Rate বাহির করিয়া cumulative f

নম্বর (1)	f (2)	cum. f (3)	cum. percent f. (4)
90—94	1	120	100·0
85—89	3	119	98·8
80—84	7	116	96·3
75—79	12	109	90·5
70—74	22	97	80·5
65—69	31	75	62·2
60—64	18	44	36·4
55—59	15	26	21·6
50—54	7	11	9·1
45—49	2	4	3·4
40—44	2	2	1·7
35—39	0	0	0·0
	N = 120		

$$\text{Rate}=\frac{1}{120}=\cdot 0083.$$

15নং টেব্‌ল। **শ্রেণীবদ্ধ সারির নম্বরের f কে cum. percent f এ পরিণত করা হইয়াছে।**

দ্বারা গুণ করিলে পাওয়া যাইবে। তবে প্রথমোক্ত প্রক্রিয়াটি সঠিক লেখচিত্রের জন্য প্রযোজ্য।

এখানে মনে রাখা দরকার যে cumulative frequency curve এর ন্যায় এখানে y অক্ষের মান (cum. percent f) সেই ধাপের উচ্চ সীমার উপর পড়িবে। যেমন 15নং টেবলে 39·5 নম্বরের উপর o, y-একক উঠিবে। 44·5 এর উপর 1·7 y-একক উঠিবে, 69·5 এর উপর 62·2 y-একক উঠিবে এবং এইরূপে শেষ ধাপে 94·5 এর উপর 100·00 y-একক উঠিবে। আর অন্যান্য বিষয় অন্যান্য লেখচিত্রানুযায়ী হইবে।

Percentile এবং percentage rank নির্ণয় ঃ

(ক) Cumulative percentage distribution হইতে PR এবং percentile বাহির করা যায়। 15নং টেবল হইতে একটি উদাহরণ লওয়া হইল। মনে করি 87 তম percentileটি বাহির করিতে হইবে। উক্ত টেবলের 1নং এবং 4নং স্তম্ভ হইতে প্রতীয়মান হইতেছে যে

74·5 পর্যন্ত 80·5%

87·0%———→… … ..————→…………→77·8

79·5 পর্যন্ত 90·5%

5·0 10%

এখানে 87 তম percentileটি 80·5% হইতে (87%−80·5%) 6·5% দূরে অবস্থিত। ∴ ত্রৈরাশিকের নিয়মে $\frac{5}{10}\times 6{\cdot}5=3{\cdot}3$ দূরত্বে অবস্থিত।

∴ 87 তম percentile হইবে 77·8 (74·5 + 3·3)।

আবার বিপরীত প্রক্রিয়াদ্বারা নম্বর হইতে PR বাহির করা যায়। উদাহরণ স্বরূপ মনে করা যাক 15নং টেবলে 57 নম্বরের PR বাহির করিতে হইবে।

54·5 পর্যন্ত 9·1%

নম্বর 57…………———→… ….. …———→15%

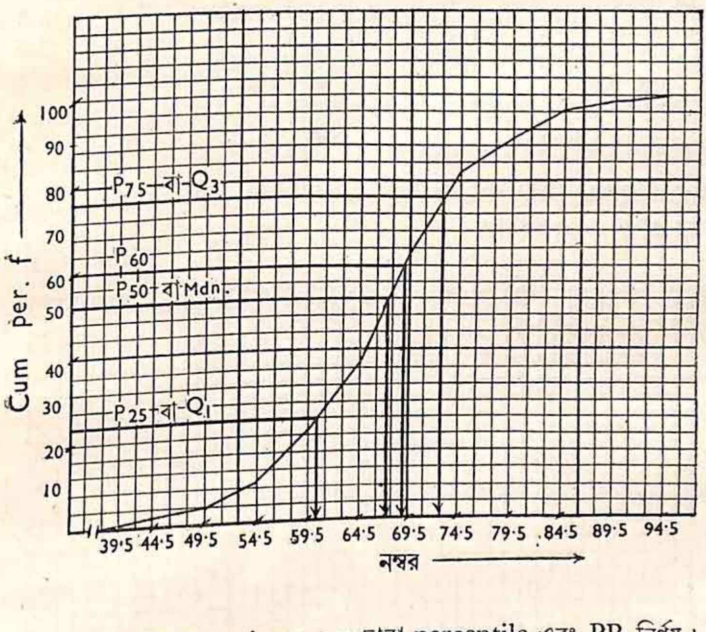
59·5 পর্যন্ত 21·6%

5·0 12·5%

পূর্বের ন্যায় স্পষ্টই দেখা যাইতেছে যে 57·0, 54·5 হইতে 2·5 নম্বর বেশী। ত্রৈরাশিকের সাহায্যে ঐ বেশী নম্বর পাওয়া যাইবে 9·1% এর পর 6·3% $\left(\frac{12{\cdot}5}{5}\times 2{\cdot}5\right)$ বেশীতে। অর্থাৎ 57·0 নম্বরের PR হইবে 9·1%+6·3% =15·4 অর্থাৎ 15 (প্রায়)।

ইহা মনে রাখিতে হইবে যে উচ্চসীমা (15নং টেবল) 84·5 এর PR 96·3 ; 74·5 এর PR 80·5. এই PR গুলি ঐ distribution এ অবস্থিত বিন্দুগুলির PR—নম্বরের PR নহে।

(খ) আবার ogive হইতে percentile এবং PR বাহির করা যায়। 6নং (A) চিত্রে বিভিন্ন PR হইতে percentile এবং বিভিন্ন percentile হইতে PR বাহির করার পদ্ধতি দেখান হইয়াছে। যেমন p_{75} নির্ণয়ে y—অক্ষের উপর 75 তম বিন্দু হইতে X-অক্ষের সমান্তরাল একটি সরলরেখা টানা হইল। এখন ঐ ছেদবিন্দু নির্দেশক নম্বরই হইল p_{75}. ঠিক এরূপভাবে অংক না কষিয়া লেখচিত্র হইতে বিভিন্ন percentile পাওয়া যায়। আবার ঠিক বিপরীত পদ্ধতিতে নম্বর হইতে PR বাহির করা যায়।



6নং (A) চিত্র :—ogive curve দ্বারা percentile এবং PR নির্ণয়।

C. **ogive এর ব্যবহার—**

(ক) **দুই বা ততোধিক দলের তুলনামূলক আলোচনার জন্য ogiveকে ব্যবহার করা হয়।** একই অক্ষদ্বয়ের উপর দুইটি ogive অঙ্কন করা যায়। নীচে 7নং চিত্রে 150 জন বালক ও 150 জন বালিকার উপর

(3) Pie লেখচিত্র :

11 নং চিত্রে কোন অঞ্চলের প্রাথমিক ; মাধ্যমিক ও টেক্নিক্যাল বিদ্যালয়ের অনুপাত দেওয়া হইল। ঐ অঞ্চলে 60% প্রাথমিক, 30% মাধ্যমিক এবং 10% টেক্নিক্যাল বিদ্যালয় আছে। 360°কে 6 : 3 : 1 ভাগে ভাগ করিয়া একটি বৃত্তকে আনুপাতিকভাবে ভাগ করা হইয়াছে।

Exercises V

1. Carry through the following steps for the first distribution of scores in Q. 2, Exercise—II.

(a) Find the cumulative frequencies, and tabulate them.

(b) Plot a cumulative distribution curve.

(c) Find the cumulative percentages and proportions, and tabulate them.

(d) Plot the ogive distribution, showing the smoothed curve.

2. The following distributions represent the achievement of two groups, A and B, upon a memory test.

(a) Plot cumulative frequency graphs of group A's and of group B's scores, observing the 75% rule.

(b) Plot ogives of the two distributions A and B upon the same axes.

(c) Determine P_{30}, P_{60} and P_{90} graphically from each of the ogives and compare graphically determined with calculated values.

(d) What is the percentile rank of score 152 in group A distribution ? In group B's distribution ?

(e) A Percentile Rank of 170 in group A corresponds to what percentile rank in group B ?

(f) What percent of group A exceeds the median of group B ?

Scores	Group A	Group B
179—183	6	8
174—178	7	8
169—173	8	9
164—168	10	16
159—163	12	20
154—158	15	18
149—153	23	19
144—148	16	11
139—143	10	13
134—138	12	8
129—133	6	7
124—128	3	2
	N=128	N=139

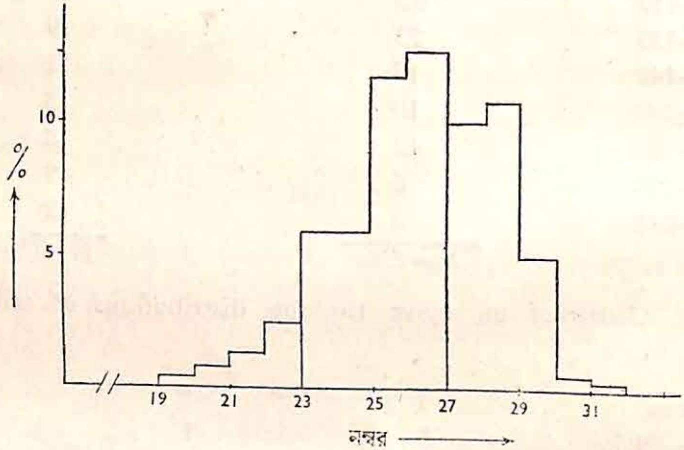
3. Construct an ogive for the distributions of scores given below.

Scores	f	f
89·5—99·5	1	1
79·5—89·5	3	0
69·5—79·5	5	2
59·5—69·5	7	9
49·5—59·5	13	15
39·5—49·5	4	8
29·5—39·5	5	9
19·5—29·5	3	3
9·5—19·5	2	6
	N=43	N=53

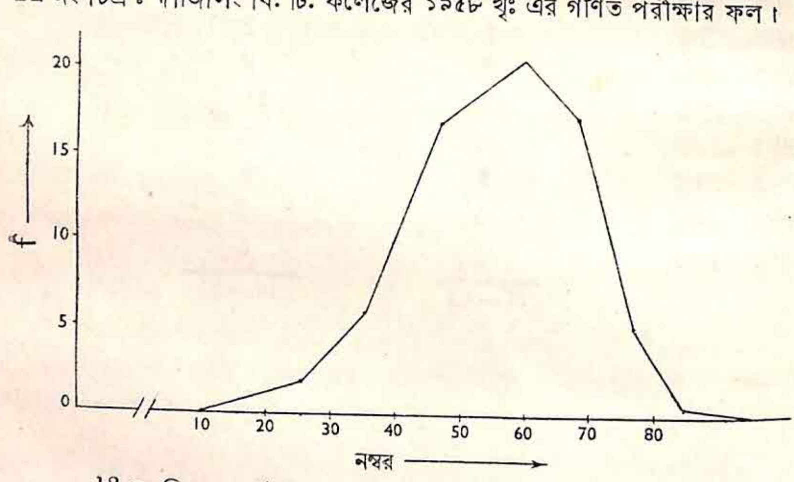
ষষ্ঠ অধ্যায়

Normal Distribution

নীচে 12 নং এবং 13 নং চিত্রদ্বয়ে সাধারণ Distribution দেখান হইয়াছে। উপর উপর মোটামুটি বুঝা যায় যে দুইটি লেখচিত্রই প্রায় সমপর্যায়-



12 নং চিত্র : দার্জিলিং বি. টি. কলেজের ১৯৫৮ খৃঃ এর গণিত পরীক্ষার ফল।

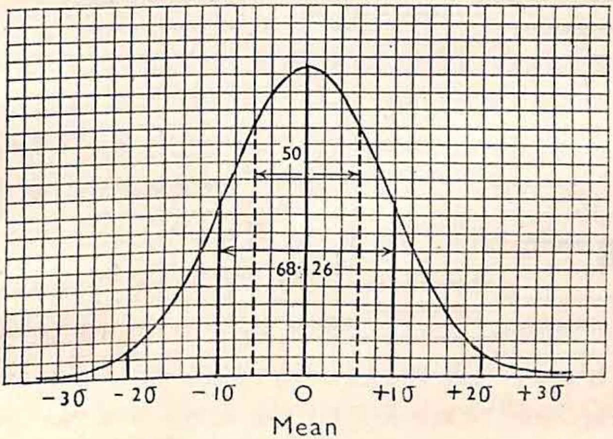


13 নং চিত্র : দার্জিলিং বি. টি. কলেজের ১৯৫৯ খৃঃ এর বিজ্ঞান পরীক্ষার ফল।

ভুক্ত। বামে বা দক্ষিণে অল্প সংখ্যক পরীক্ষার্থী বিদ্যমান এবং মাঝখানে অধিকাংশ পরীক্ষার্থীর ভিড়। নিম্নস্তরে কমসংখ্যক পরীক্ষার্থী। পরে

মাঝখানে পরীক্ষার্থীর ভিড় এবং তার জন্য সৃষ্টি করিয়াছে শৃঙ্গ। শৃঙ্গের পরে ক্রমশঃ ঢালু হইয়া ভূমির সহিত মিশিয়াছে।

উপরোক্ত দুই চিত্রের শৃঙ্গদ্বয় হইতে যদি ভূমির উপর লম্ব অঙ্কিত করা যায় তাহা হইলে প্রতিটি লেখচিত্র এমন দুইটি ভাগে বিভক্ত হইবে যে দুইটি অংশ প্রায় সদৃশ এবং সেইজন্য দুই বিভাগের ক্ষেত্রও প্রায় সমান। নীচে এমন একটি লেখচিত্র প্রদত্ত হইল যাহার পূর্ববর্ণিত শৃঙ্গদ্বয় সদৃশ এবং দুই পাশের ক্ষেত্রফলও নির্ভুলভাবে সমান। এরূপ উভয়পার্শ্বসদৃশ ঘণ্টাকৃতি চিত্রকে Normal Probability curve বা সংক্ষেপে Normal curve বলে। এই অধ্যায়ে Normal probability curve সম্বন্ধে আলোচনা করা হইবে।



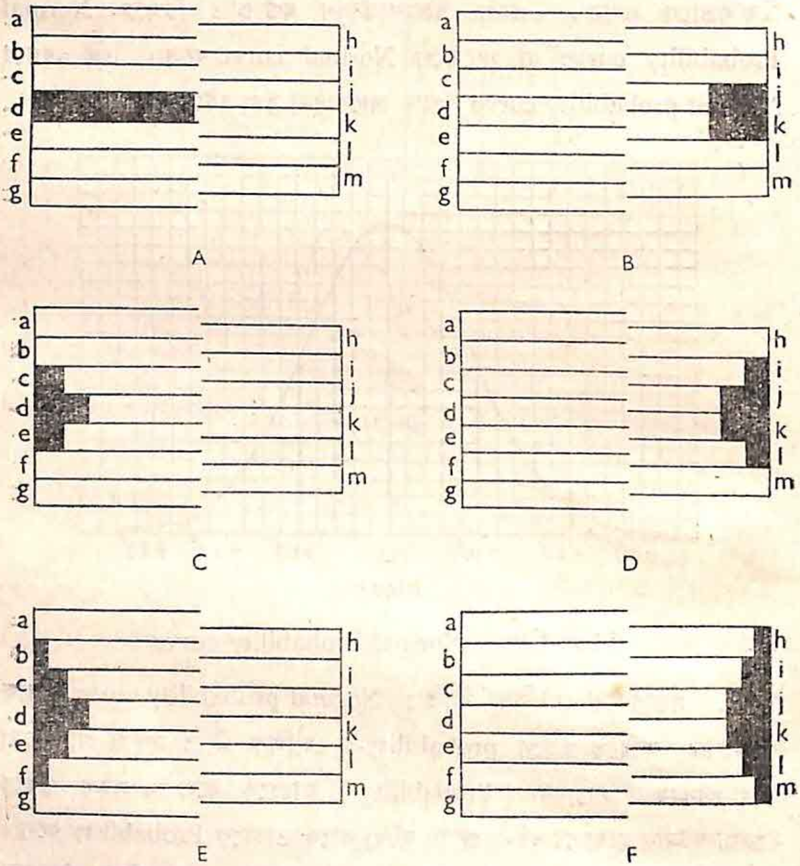
14 নং চিত্র। Normal Probability curve

1. সম্ভাবনার মৌলিক নীতি : Normal probability curve সম্বন্ধে আলোচনা করিতে হইলে probability-র মৌলিক নীতি সম্বন্ধে আলোচনা করা দরকার। সম্ভাবনা (Probability) কাহাকে বলে? সকল প্রকার সম্ভাব্য ঘটনার মধ্যে যে ঘটনা প্রায়ই ঘটিয়া থাকে তাহাকে Probability বলে। সাধারণতঃ পাশার ঘুঁটি ফেলা, মুদ্রা ছোঁড়া প্রভৃতির দ্বারা ঘটনার সম্ভাবনা সম্বন্ধে জ্ঞান লাভ করা যাইতে পারে।

কোন ঘটনার সম্ভাবনাকে গণিতে অনুপাত দ্বারা প্রকাশ করা যাইতে পারে। নিরপেক্ষভাবে বিচার করিলে দেখা যায় যে ছুঁড়িয়া দিলে একটি মুদ্রার মাথার দিকের পড়িবার সম্ভাবনা আছে 1/2 এবং পাশার ছক্কার পড়িবার সম্ভাবনা আছে

1/6. এই অনুপাতগুলির নাম Probability—তাহাদের হর হইবে মোট সম্ভাবনা এবং লব হইবে অভীপ্রিত সম্ভাবনা। তাই ছক্কার যে কোন একটি দিকের Probability হইবে 1/6 অথবা অভীপ্রিত ঘটনা—মোট সম্ভাবনা। এই সকল অনুপাত 0·00 হইতে 1·00 এর মধ্যে অবস্থিত।

15 নং চিত্র Normal Probability curve এর সূত্র সম্বন্ধে জ্ঞান দানে সাহায্য করিবে। দুইটি একমুখ খোলা ঘরের সারি চিত্রের ন্যায় আঁটা আছে।



15নং চিত্র Normal Probability curve পাওয়ার একটি পদ্ধতি।

d ঘরে (15নং চিত্র—A) কিছু বালি বা ময়দা জাতীয় গুঁড়া পদার্থ রাখা হইল। এখন উল্লিখিত যন্ত্রের বাম দিক যদি কিছু উঁচু করা যায় এবং ঈষৎ টোকা দেওয়া যায়, দেখা যাইবে যে ঐ বালি গড়াইয়া ডানদিকের j ঘরে অর্ধেক এবং k ঘরে অর্ধেক পড়িবে (চিত্র B)। ইহাকে $\frac{1}{2}+\frac{1}{2}$ Binomial বলিয়া

অভিহিত করা হয়। এখন B চিত্রের দক্ষিণ দিকটি যদি উঁচু করা হয় তবে দেখা যাইবে যে j ঘরের $\frac{1}{2}$ বালি c ঘরে, k ঘরের $\frac{1}{2}$ বালি e ঘরে এবং j ও k ঘরের প্রত্যেকের অর্ধেক বালি d ঘরে গড়াইয়া পড়িয়াছে (c চিত্র)। তবেই দেখা গেল মোট বালির $\frac{1}{4}$ অংশ c ঘরে, $\frac{1}{4}$ অংশ e ঘরে এবং $\frac{1}{2}$ d ঘরে স্থান পাইয়াছে। ইহাকে Binomial $(\frac{1}{2}+\frac{1}{2})^2$ নামে আখ্যা দেওয়া হইয়া থাকে। এখন বাম দিকটি আবার উঁচু করিলে c হইতে অর্ধেক i ঘরে এবং অর্ধেক j ঘরে যাইবে। d হইতে অর্ধেক j এবং বাকী অর্ধেক k ঘরে যাইবে। e হইতে অর্ধেক k এবং বাকী অর্ধেক l ঘরে যাইবে (চিত্র D)। ফলে সমস্ত বালির $\frac{1}{8}$ অংশ i, $\frac{3}{8}$ অংশ j, $\frac{3}{8}$ অংশ k এবং $\frac{1}{8}$ অংশ l ঘরে রহিল। ইহাকে binomial $(\frac{1}{2}+\frac{1}{2})^3$ দ্বারা প্রকাশ করা হয়।

এইরূপে আবার দক্ষিণ দিক উঁচু করিলে মোট বালির $\frac{1}{16}$ l ঘরে, $\frac{4}{16}$ c ঘরে $\frac{6}{16}$ d ঘরে, $\frac{4}{16}$ e ঘরে, এবং $\frac{1}{16}$ f ঘরে যাইবে (চিত্র e)। ইহা binomial এ $(\frac{1}{2}+\frac{1}{2})^4$ দ্বারা স্থচিত হয়। পরে আবার বাম দিকে উঁচু করিলে মোট বালির $\frac{1}{32}$ অংশ h ঘরে, $\frac{5}{32}$ অংশ i ঘরে, $\frac{10}{32}$ অংশ j ঘরে, $\frac{10}{32}$ অংশ k ঘরে, $\frac{5}{32}$ অংশ l ঘরে এবং $\frac{1}{32}$ অংশ m ঘরে গড়াইয়া পড়িবে (চিত্র f) ইহাকে binomial $(\frac{1}{2}+\frac{1}{2})^5$ দ্বারা প্রকাশ করা হইয়া থাকে।

এইভাবে যদি আমরা আরও অগ্রসর হই তবে binomial এর দ্বারা আমাদের সম্ভাবনাগুলিকে দেখাইতে পারিব। কিন্তু এই পদ্ধতি সময় ও কষ্ট সাপেক্ষ। তাই আমরা মুদ্রা ছোঁড়ার মাধ্যমে ঐ স্থত্রে আসিবার চেষ্টা করিব। সেখানে আমরা এমন সব মুদ্রা ব্যবহার করি যাহাদের মাথা ও লেজ—দুইদিকের পড়ার সম্ভাবনা সমভাবে বিদ্যমান এবং তাহারা কোনক্রমে কিনারার উপর দাঁড়াইতে পারে না। এইরূপ একটি মুদ্রার মাথা (Head) এবং লেজ (Tail) পড়ার মোট সম্ভাবনাকে $\frac{1}{2}H+\frac{1}{2}T$ দ্বারা প্রকাশ করা হয়। যদি দুইটি মুদ্রা ছুঁড়িয়া দিই তবে নিম্নরূপ সম্ভাবনা স্বাভাবিক—

1	2	3	4
a b	a b	a b	a b
H H	H T	T H	T T

প্রথমটিতে দুইটি মুদ্রার মাথা (H) উপরে থাকিতে পারে। দ্বিতীয়টিতে প্রথমটি মাথা (H) এবং অন্যটির লেজ (T) উপরে থাকিতে পারে। তৃতীয়টিতে

দ্বিতীয়টির উল্টা সম্ভাবনা এবং চতুর্থটিতে দুইটির লেজই (T) উপরে থাকিতে পারে। উপরের সম্ভাবনাগুলিকে নিম্ন সূত্রাকারে লেখা যায়।

$$(H+T)^2 = H^2 + 2\,HT + T^2$$

অর্থাৎ $1H^2$—দুইটি H হইবার 4এ 1টি সম্ভাবনা। ∴ সম্ভাবনার হার $\frac{1}{4}$

$2HT$—একটি H এবং একটি T হইবার 4এ দুটি সম্ভাবনা। ∴ সম্ভাবনার হার $\frac{2}{4}$

$1T^2$—দুইটি T হইবার 4এ 1টি সম্ভাবনা ∴ সম্ভাবনার হার $=\frac{1}{4}$

মোট সম্ভাবনা = 4টি

আবার আমরা যদি তিনটি মুদ্রা টস্ করি তবে নিম্নরূপ ঘটনা সম্ভবপর।

1	2	3	4	5	6	7	8
a b c	a b c	a b c	a b c	a b c	a b c	a b c	a b c
HHH	HHT	HTH	THH	TTH	THT	HTT	TTT

অর্থাৎ তিনটি মুদ্রা টসের সম্ভাবনাকে নিম্নরূপ সূত্রে প্রকাশ করা যায়।

$$(H+T)^3 = H^3 + 3\,H^2T + 3\,HT^2 + T^3$$

অর্থাৎ $1H^3$ তিনটি H হইবার সম্ভাবনা 8টির মধ্যে 1টি ∴ সম্ভাবনার হার $=\frac{1}{8}$

$3H^3T$ দুইটি H এবং একটি T হইবার সম্ভাবনা 8টির মধ্যে 3টি ∴ ,, ,, $=\frac{3}{8}$

$3HT^2$ একটি H এবং দুইটি T ,, ,, 8টির ,, 3টি ∴ ,, ,, $=\frac{3}{8}$

$1T^3$ তিনটি T হইবার সম্ভাবনা 8টির মধ্যে 1টি ∴ ,, ,, $=\frac{1}{8}$

মোট সম্ভাবনা = 8টি

অতএব উল্লিখিত টস্ হইতে এই সিদ্ধান্তে উপনীত হওয়া যায় যে টসের সম্ভাবনাকে আমরা binomial সূত্রদ্বারা প্রকাশ করিতে পারি। যেমন 6টি মুদ্রাকে টস্ করিলে।

$$(H+T)^6 = H^6 + 6H^5T + 15H^4T^2 + 20H^3T^3 + 15H^2T^4 + 6HT^5 + T^6.$$

এখানে H^6 এর সম্ভাবনার হার $=\frac{1}{64}$

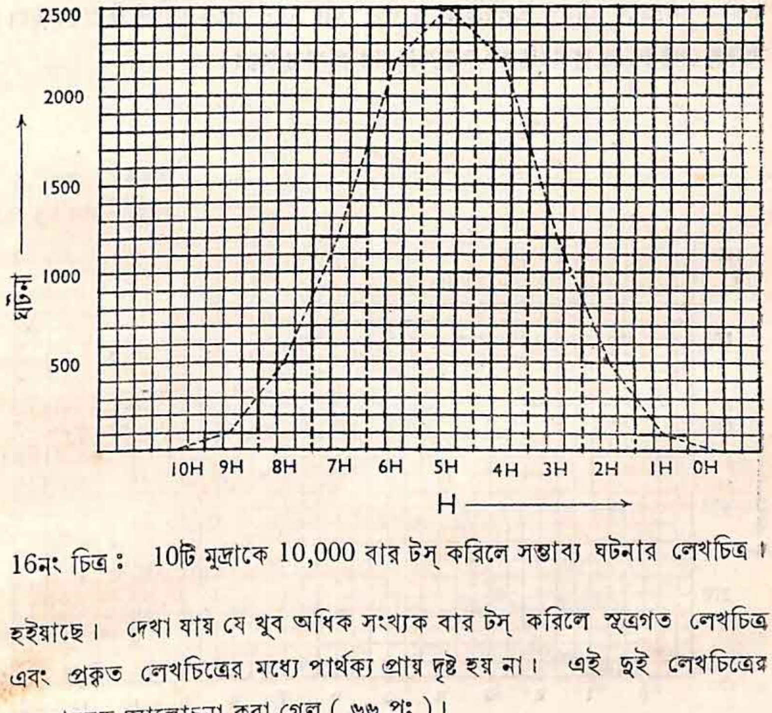
H^5T	,,	,,	,,	$=\frac{6}{64}$
H^4T^2	,,	,,	,,	$=\frac{15}{64}$
H^3T^4	,,	,,	,,	$=\frac{20}{64}$
H^2T^4	,,	,,	,,	$=\frac{15}{64}$
HT^5	,,	,,	,,	$=\frac{6}{64}$
T^6	,,	,,	,,	$=\frac{1}{64}$

সেইরূপ 10টি মুদ্রাকে টস্ করিলে

$$(H+T)^{10} = H^{10} + 10H^9T + 45H^8T^2 + 120H^7T^3 + 210H^6T^4 + 252H^5T^5 + 210H^4T^6 + 120H^3T^7 + 45H^2T^8 + 10HT^9 + T^{10}.$$

শেষোক্ত ফলটিকে লেখচিত্রের সাহায্যে 16নং চিত্রে দেখান হইল।

এইভাবে মুদ্রা বা ছক্কা লইয়া টস্ করিয়া বিভিন্ন পরীক্ষা নিরীক্ষা হইয়াছে। বহুবার টস্ করিয়া সূত্রগত লেখচিত্রের সহিত প্রকৃত লেখচিত্রের তুলনা করা



16নং চিত্র : 10টি মুদ্রাকে 10,000 বার টস্ করিলে সম্ভাব্য ঘটনার লেখচিত্র।

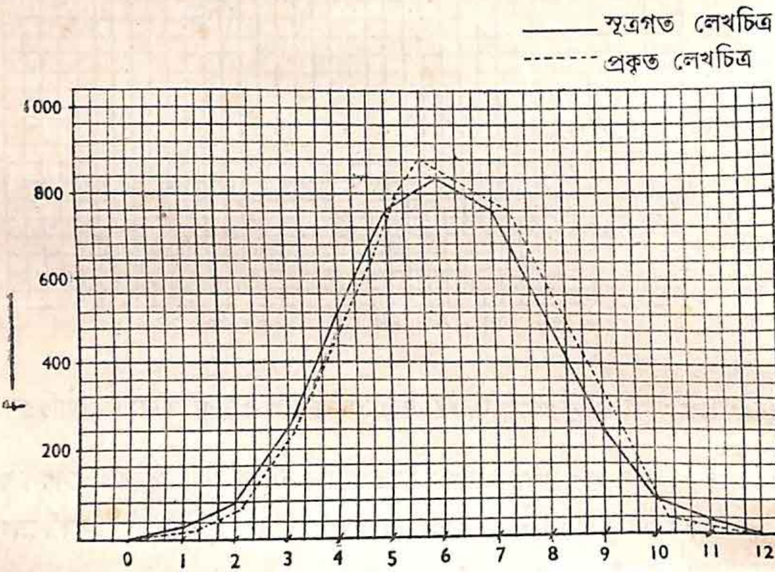
হইয়াছে। দেখা যায় যে খুব অধিক সংখ্যক বার টস্ করিলে সূত্রগত লেখচিত্র এবং প্রকৃত লেখচিত্রের মধ্যে পার্থক্য প্রায় দৃষ্ট হয় না। এই দুই লেখচিত্রের তুলনামূলক আলোচনা করা গেল (৬৬ পৃঃ)।

এখানে মনে রাখা দরকার যে টস্ প্রক্রিয়ায় ঘটনার সম্ভাবনা এবং সম্ভাবনা হীনতা পূর্ববর্ণিত মুদ্রার ন্যায় সমান নাও হইতে পারে ; কিন্তু তাহারা যে কোন অনুপাতে (যেমন $\frac{1}{3}H+\frac{2}{3}T$) থাকিলেও তাহাদের যোগফল সর্বদা 1·00 হইবে। যখন H এবং T সমান নয় তখন $(H+T)^n$ এর বিস্তৃতির লেখচিত্র Normal হইবে না।

Probability-র ব্যবহার

16নং চিত্রে $(H+T)^{10}$ এর বিস্তৃতিকে লেখচিত্রে প্রকাশ করা হইয়াছে। যদি 10টি মুদ্রার পরিবর্তে 20, 30 অথবা 100টি মুদ্রা লওয়া হইত তাহা হইলে ঐ Polygon curve এর বাহুর সংখ্যা বৃদ্ধি পাইত এবং দৈর্ঘ্যও কমিয়া যাইত।

এইরূপে Polygon এর কৌণিক বিন্দুগুলি ক্রমে নিকটে আসিতে থাকে। যদি $(H+T)^n$ এর বিস্তৃতিতে n অগণিত হয়, তখন ঐ বিন্দুগুলি পাশাপাশি বসিয়া একটি রেখা দ্বারা Normal Probability curve প্রদান করে (14নং চিত্র)। বাস্তবে ঐরূপ লেখচিত্র পাওয়া যায় না; তবে আদর্শ হিসাবে ইহা বিভিন্ন লেখচিত্রের তুলনামূলক আলোচনার সাহায্য করে।



17নং চিত্র। 12টি পাশাকে 4096 বার টস্ করিয়া সূত্রগত ও প্রকৃত লেখচিত্রের তুলনা মূলক অবস্থান। (Yule হইতে)

12নং এবং 13নং চিত্রের দিকে বিশেষ ভাবে দৃষ্টি দিলে স্পষ্টই প্রতীয়মান হয় যে উহাদের সহিত 14নং চিত্রের কিছু সাদৃশ্য বিদ্যমান। এই ঘণ্টাকৃতি (bell shape) লেখচিত্র সমূহের নীতিকে এই বলিয়া ব্যাখ্যা করা হয় যে প্রাকৃতিক বহু অবস্থা এবং মানসিক ও সামাজিক গুণগুলির পরিমাপের সময় মধ্যস্থলে স্ফীত হওয়ার ঝোঁক বিদ্যমান এবং প্রায় Normal curve-এর লেখচিত্রে উহারা পরিদৃশ্যমান হয়। কি জীববিজ্ঞান—কি সমাজ বা অর্থনীতি বিজ্ঞান—কি মনোবিজ্ঞান—এমন কি মানুষের ভ্রান্তির পরিমাপ প্রভৃতি ক্ষেত্রও Normal curve-এর উপযোগিতা উপলব্ধি করিয়া থাকে।

Normal Probability curve-এর ধর্ম

(ক) Normal curve-এর সমীকরণ নীচে প্রদত্ত হইল।

$$Y = \frac{N}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}} \qquad \text{.........................(20)}$$

যখন x = Mean হইতে কোন নম্বরের দূরত্ব। ইহা ভূমির উপর অবস্থিত।

এবং Y = X — অক্ষের উপর লেখচিত্রের উচ্চতা।

সূত্রের অন্যান্য পদগুলি ধ্রুবক।

যেমন N = মোট সংখ্যা

σ = শ্রেণীবদ্ধ সারির SD

$\pi = 3{\cdot}1416$

এবং $e = 2{\cdot}7183$

যখন N এবং σ দেওয়া থাকে, তখন 20নং সূত্র দ্বারা (i) প্রদত্ত x এর উচ্চতা বা Y এবং (ii) দুইটি মাননির্দেশক বিন্দুদ্বয়ের মাঝখানে ও কোন প্রদত্ত বিন্দুর উপরে বা নীচে শতকরা কতজন বিদ্যমান—তাহা বাহির করা যায়। কিন্তু প্রতিবার এই দীর্ঘ গণনার সহায়তা না লইয়া আমরা A টেবলের দ্বারস্থ হই। A টেবল অপরিহার্য বলিয়া আমরা নীচে কিছু আলোচনা করিব।

(খ) Normal curve-এর মধ্যে অবস্থিত ক্ষেত্রের পরিমাপ :—A টেবলে Normal curve-এ Mean এবং X-অক্ষের উপর বিভিন্ন বিন্দুর উপর অংকিত লম্বের মধ্যে অবস্থিত ক্ষেত্র সমূহ, মোট লেখচিত্র—আবদ্ধ ক্ষেত্রের কত কত ভগ্নাংশ—তাহার নির্দেশ দেয়। Normal curve-এর মোট ক্ষেত্রকে 10,000 একক ধরিয়া ভিন্ন অংশকে সেই অনুপাতে প্রকাশ করা হইয়া থাকে।

A টেবলের প্রথম সারি $\frac{x}{\sigma}$ দ্বারা সূচিত হইয়াছে। Mean হইতে আরম্ভ করিয়া বিভিন্ন দূরত্বে $\frac{x}{\sigma}$ এর মান কত তাহা ঐ সারিতে দেওয়া হইয়াছে। পূর্বে আলোচিত হইয়াছে যে $X - M = x$. Mean হইতে কোন নম্বরের দূরত্বকে σ তে প্রকাশ করিলেই $\frac{x}{\sigma}$ পাওয়া যায়। ইহাকে এক কথায় sigma-score বা Z-score বলা হইয়া থাকে। Mean হইতে $\cdot 5\sigma$ দূরে কি সংখ্যা বিদ্যমান তাহা বাহির করিতে হইলে A টেবলে x/σ সারির উপর হইতে ·5

ধাপ পর্য্যন্ত নামিতে হইবে। এখন উক্ত ·5 বরাবর ডান দিকে পরবর্ত্তী স্তম্ভে ·00-এর নীচে 1915 বিদ্যমান। এই সংখ্যা দ্বারা এই বুঝায় যে মোট লেখচিত্রের ক্ষেত্রফল 10,000 এর মধ্যে 1915 জন ঐ ক্ষেত্রে বিদ্যমান। অর্থাৎ Mean এবং $.5\sigma$ এর মধ্যে 19.15% জন আছে। এইরূপে Mean এবং $1{\cdot}23\sigma$ এর মধ্যে অবস্থিত লোক সংখ্যা নির্ণয় করিতে হইলে A টেবলে প্রথম স্তম্ভে 1·2 পর্য্যন্ত নামিয়া সেই বরাবর দক্ষিণে অবস্থিত ·03 স্তম্ভের সোজা নীচের সংখ্যা অর্থাৎ 3907 কে ধরিতে হইবে। অর্থাৎ Normal curve এ 39·07% লোক M এবং $1{\cdot}23\sigma$ এর মধ্যে অবস্থিত। এখানে স্মরণ রাখা উচিত যে লেখচিত্রটি Mean এর দুই পাশে সদৃশ বলিয়া '+' এবং '—' দূরত্বের ক্ষেত্রফল সমান।

Normal curveটি কখনও × অক্ষের সহিত মিলিত হয় না। তবে ব্যবহারিক ক্ষেত্রে আমরা $+3\sigma$ এবং -3σ দূরত্ব দ্বারা লেখচিত্রের ক্ষেত্রটি সীমাবদ্ধ করি। $M \pm 3\sigma$ মধ্যবর্ত্তী স্থানে 99·73% লোক বর্ত্তমান থাকায় ঐ সীমা দ্বারাই আমরা প্রায় 100 জনের অবস্থান নির্দেশ করি। তাহার জন্য আমরা যে ·27% ভুলের সম্মুখীন হই তাহা তত মারাত্মক ভুল নহে। তবুও সর্বদা আমরা ঐ বিষয়ে সচেতন থাকিব।

(গ) Normal Probability curve-এর অন্যান্য ধ্রুবক সম্বন্ধে আলোচনা ঃ Normal curve হইতে ইহা প্রতীয়মান হয় যে, M, Mdn এবং Mode লেখচিত্রের মধ্যস্থলে একই বিন্দুতে অবস্থিত থাকে। Normal curve-এ $M \pm \sigma$-এর সীমার মধ্যে মাঝের 68·26% লোক থাকে ; $M \pm 2\sigma$-এর সীমার মধ্যে মাঝের প্রায় 95% লোক থাকে ; এবং $M \pm 3\sigma$-এর সীমার মধ্যে 99·7% লোক বর্ত্তমান থাকে। ভিন্ন ভাষায় বলা যায় যে Normal distribution-এ কোন ব্যক্তি যে Mean হইতে $\pm\sigma$ দূরত্বের মধ্যে থাকিবে তাহার সম্ভাবনা আছে শতকরা 68 ; এইরূপে কোন ব্যক্তি Mean হইতে $\pm 2\sigma$ দূরত্বের মধ্যে থাকিবে —তাহার সম্ভাবনা আছে শতকরা 95 ; এবং শতকরা 99·7 সম্ভাবনা আছে কোন ব্যক্তি Mean হইতে $\pm 3\sigma$ দূরত্বের মধ্যে থাকিবে।

আবার σ এর বদলে Q দ্বারা দূরত্ব প্রকাশ করা যাইতে পারে। Normal curve-এ Q কে Probable Error বা PE বলা হইয়া থাকে।

$$PE = {\cdot}6745\ \sigma$$

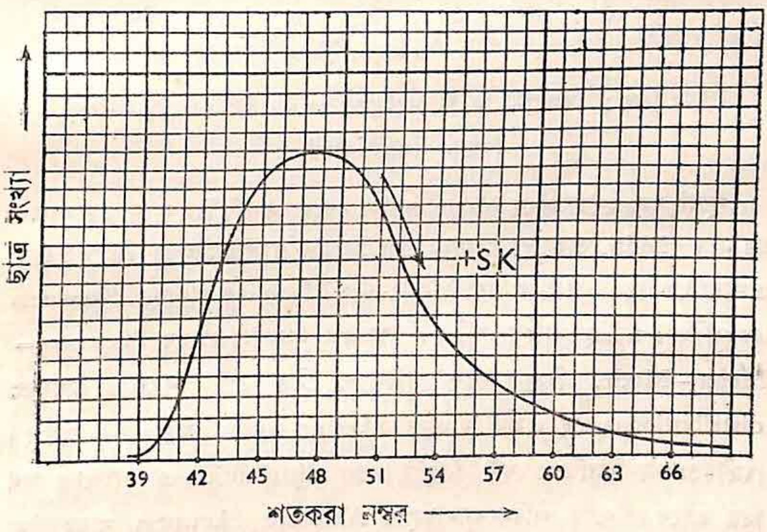
অর্থাৎ $\sigma = 1{\cdot}4826$ PE.

A টেবল হইতে হিসাব করিয়া বলা যায় যে M±PE এর মধ্যে মাঝের 50%, M±2PE-এর মধ্যে মাঝের 82·26%, M±3PE এর মধ্যে মাঝের 97·7% এবং M±4PE এর মধ্যে মাঝের শতকরা 99·30% জন লোক বিদ্যমান।

B. Non-normal Distribution

যে distributionগুলি Normal নয় তাহাদিগকে Non-Normal distribution বলে। সাধারণতঃ আমরা কার্যক্ষেত্রে Normal curve পাই না। Non-Normality-র একটি লক্ষণ হইল (ক) Skewness এবং অপরটি (খ) Kurtosis.

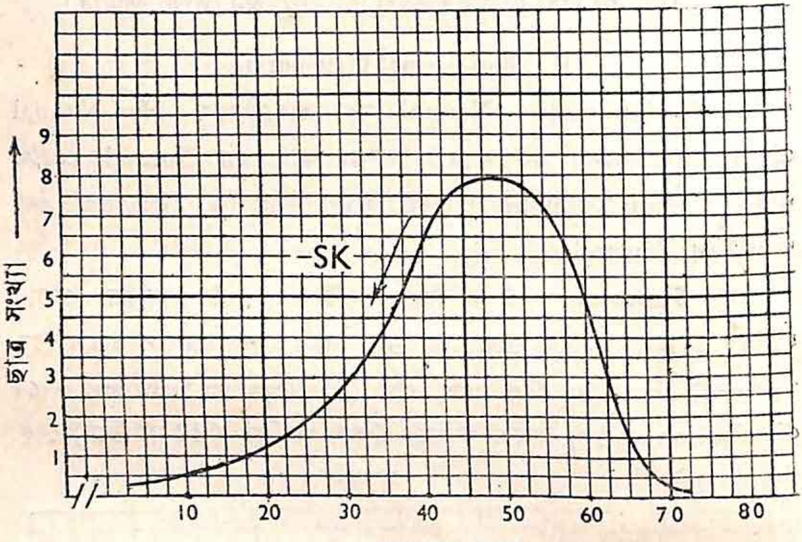
(ক) **Skewness** : 12 নং চিত্রে একটি Skewed লেখচিত্র দেখান হইয়াছে। ঢাল হিসাবে লেখচিত্রের দুই রকম Skewness পাওয়া যাইতে পারে—(i) Positive Skewness এবং (ii) Negative Skewness। যে distribution এর ঢালু লেজ লেখচিত্রের দক্ষিণে অবস্থিত (18 নং চিত্র) সেই



18 নং চিত্র : কল্যাণী বি. টি. কলেজের ১৯৬০ খ্রীঃ-এর টেষ্ট পরীক্ষায় গণিত পদ্ধতির নম্বর।

distribution-টির Positive Skewness আছে। যেমন—মাহিনা, ইলেক্‌ট্রিক খরচ—বয়স্ক পুরুষের ওজন প্রভৃতির distribution এই শ্রেণীভুক্ত।

আবার লেখচিত্রের ঢালু লেজ যদি বামদিকে থাকে (19 নং চিত্র) তাহা হইলে সেই distribution-টির Negative Skewness থাকিবে।



19 নং চিত্র : কল্যাণী বি. টি. কলেজের ১৯৬০ খ্রীঃ-এর টেষ্ট পরীক্ষায় বিজ্ঞান পদ্ধতির নম্বর।

পূর্বেই বলা হইয়াছে যে Mode প্রান্তীয় নম্বরের উপস্থিতির দ্বারা প্রভাবান্বিত হয় না। Mdn. তাহাদের অবস্থান এবং Mean প্রান্তীয় নম্বরের গুরুত্বের দ্বারা প্রভাবান্বিত হয়। তাই আমরা Mode এবং Mean এর সাহায্যে Skewness এর পরিমাণ ও চিহ্ন বাহির করিব। সুতরাং দাঁড়াইল এই যে, Skewness = Mean − Mode. কিন্তু এরূপ পরিমাপে কিছু ত্রুটি আছে। প্রথমতঃ distribution-এ যে একক ব্যবহৃত Skewness-এরও সেই একক হইবে। দ্বিতীয়তঃ অল্প বিস্তৃতি ও বেশী বিস্তৃতি বহুল distribution-এ তাহাদের অর্থ ভিন্ন হইবে। তাই পরিসংখ্যানবিদ্‌গণ Absolute Skewness ব্যবহার না করিয়া relative skewness ব্যবহার করেন এবং উক্ত Skewness-কে σ দ্বারা ভাগ করিয়া তাহা সম্ভব করেন। তাহা হইলে সূত্র হইল

$$\text{Skewness} = \frac{\text{Mean} - \text{Mode}}{\sigma} \quad \ldots\ldots\ldots\ldots\ldots\ldots (21)$$

আবার Modeও নির্ভরযোগ্য পরিমাপ নয় বলিয়া Mode এর পরিবর্তে

Mdn কে সূত্রে স্থান দেওয়া হইয়া থাকে। Karl Pearson বাস্তব অভিজ্ঞতা দ্বারা এই সিদ্ধান্তে আসেন যে Mode হইতে Mdn. এর দূরত্ব Mean হইতে Median এর দূরত্বের দ্বিগুণ। তাই

$$\text{Skewness} = \frac{3(\text{Mean}-\text{Median})}{\sigma} \quad \ldots\ldots\ldots\ldots(22)$$

এখন 11 নং টেবলের distribution-এর Skewness বাহির করিলে 22 নং সূত্রানুযায়ী Skewness $= -\cdot 1$ [কারণ উহার $M = 30\cdot 80$, $Mdn = 32\cdot 00$ এবং $\sigma = 12\cdot 63$]। সুতরাং ঐ শ্রেণীগুচ্ছের Negative Skewness বিদ্যমান; Normal হইলে Skewness 0 হইত।

22 নং সূত্র ছাড়া percentile-এর সাহায্যে Skewness বাহির করা যায়। নীচে সূত্র দেওয়া হইল।

$$SK = \frac{P_{90}+P_{10}}{2} - P_{50} \quad \ldots\ldots\ldots\ldots(23)$$

এখানে মনে রাখা দরকার যে 22 নং সূত্র এবং 23 নং সূত্রের দ্বারা যে Skewness বাহির হইবে তাহাদের মধ্যে কিছু তফাৎ থাকা স্বাভাবিক। Normal-এর কাছাকাছি হইলে ঐ দুই প্রকারে নির্ণীত Skewness 0 এর কাছাকাছি থাকিবে। তবে দুইটি distribution-এর Skewness-এর তুলনা করিবার সময় উভয় ক্ষেত্রে একই সূত্র ব্যবহার করা উচিত।

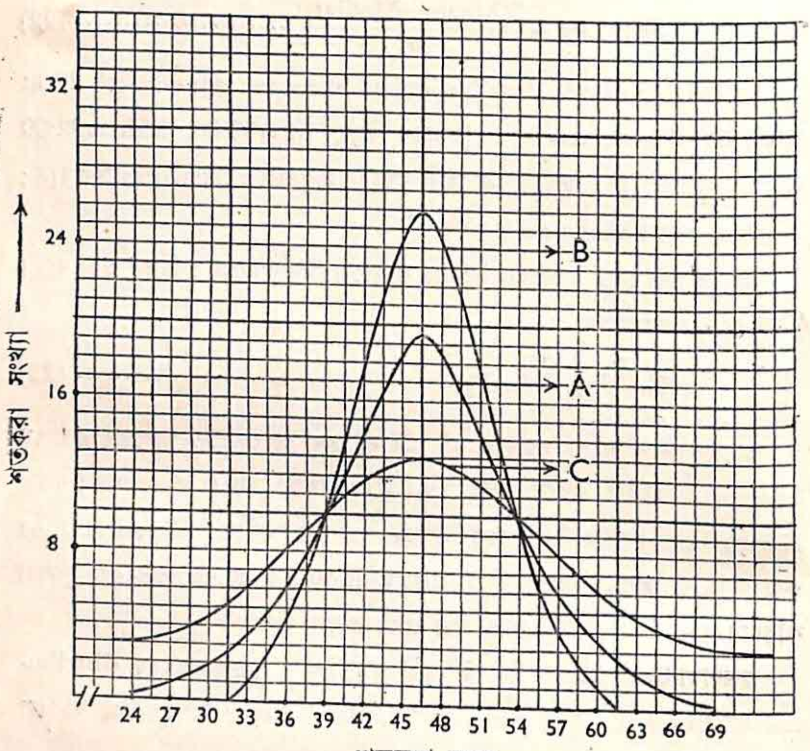
সূত্রদ্বারা Skewness বাহির করিয়াই ঘোষণা করা উচিত নয় যে distribution-টি Skewed. Skewness-এর Standard Error বাহির করিয়া সে ঘোষণা সম্ভবপর। অষ্টম অধ্যায়ে তাহা আলোচিত হইবে।

(খ) **Kurtosis** Skewness ছাড়া Non-Normality-র আর একটি লক্ষণ হইল kurtosis. Normal curve অপেক্ষা কোন curve-এর আকৃতি শঙ্কু বা চ্যাপ্টা হইতে পারে। Normal curve-এর তুলনায় শঙ্কু হইলে Leptokurtic এবং চ্যাপ্টা হইলে Platykurtic বলা হইয়া থাকে। Normal curve-কে Mesokurtic বলে। নীচে 20নং চিত্রের সাহায্যে দুই রকম Non-Normal curve-এর সংগে Normal curve-এর তুলনা করা হইয়াছে।

কোন distribution-এ kurtosis আছে কি না বা কতটা আছে তাহা বিভিন্ন সূত্র দ্বারা বাহির করা হয়। নীচে একটি সূত্র দেওয়া হইল।

$$\text{Kurtosis বা ku} = \frac{Q}{P_{90}-P_{10}} \quad \ldots\ldots\ldots\ldots(24)$$

Normal curve-এ এই সূত্রে Ku-এর মান হইবে ·263. যদি Ku ·263-এর বেশী হয় তাহা হইলে distribution-টি Platykurtic এবং কম হইলে Leptokurtic হইবে।



20নং চিত্র : কল্যাণী বি. টি. কলেজের ১৯৬০ খ্রীঃ এর টেষ্ট পরীক্ষায় ছাত্র ও ছাত্রীদের নম্বরের লেখচিত্র। B লেখচিত্রটি ছাত্রীদের এবং C লেখচিত্রটি ছাত্রদের।

A. Mesokurtic. ; B. Leptokurtic. ; C. Platykurtic.

এখন 11নং টেবলের distribution-টির Ku বাহির করা গেল। Q= 8·28; $P_{90}=47{\cdot}00$, এবং $P_{10}=12{\cdot}00$.

$$\therefore \quad Ku=\frac{8{\cdot}28}{47{\cdot}00-12{\cdot}00}= \cdot 236.$$

যেহেতু উপরোক্ত Ku=·236 অর্থাৎ ·263 অপেক্ষা কম, সুতরাং 11নং টেবলের distribution-টি Leptokurtic. কোন distribution Normal

কি না তাহা পরীক্ষা করিতে হইলে তাহার Skewness এবং kurtosis দুইই বাহির করিতে হইবে। বলা বাহুল্য Skewness-এর ন্যায় kurtosis-এরও Standard Error বাহির করিয়া তাহার যাথার্থতা যাচাই করা হয়।

(C) একই ক্ষেত্র, Mean এবং σ বিশিষ্ট Histogram বা Frequency Polygon-এর সহিত Normal curve-এর তুলনা কি করিয়া সম্ভব বা কি করিয়া ঐ Histogram বা Polygon-কে Normal curve দ্বারা "best fitting' সম্ভব তাহা এখন আলোচিত হইবে। যদি কোন distribution-এর ক্ষেত্র (N) এবং বিস্তৃতি (σ) দেওয়া থাকে তবে "best fit" Normal curve-এর অংকন পদ্ধতি নীচে 17নং টেবলে দেওয়া হইল।

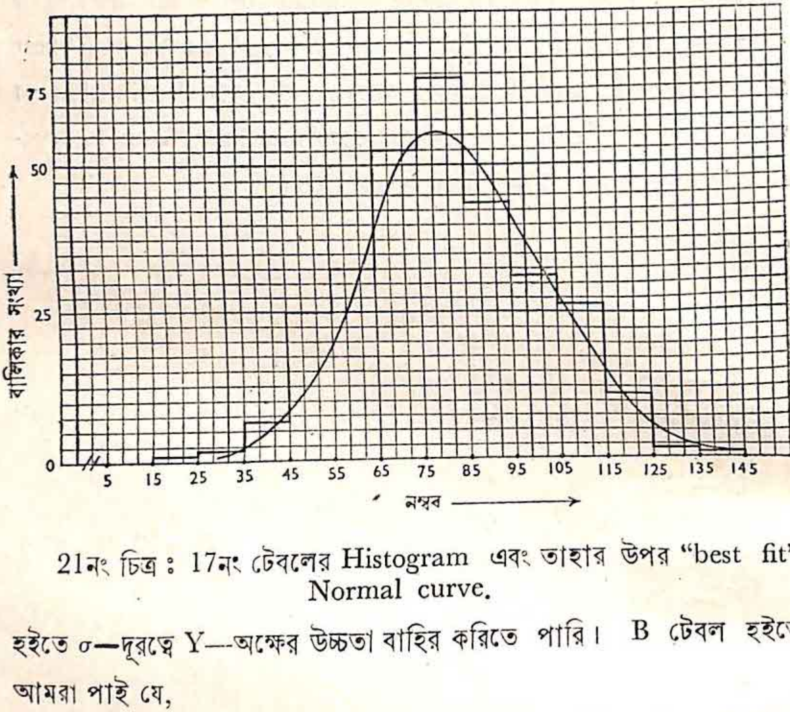
	f
135—144	1
125—134	4
115—124	11
105—114	27
95—104	31
85— 94	44
75— 84	64
65— 74	53
55— 64	33
45— 54	25
35— 44	7
25— 34	2
15— 24	1
	N = 303

N = 303
Mean = 80·63
σ = 20·95 = 2·1 ধাপ-একক
$\sqrt{2\pi}$ = 2·51.

17নং টেব্‌ল : কলিকাতার কোন বিদ্যালয়ে বালিকাদের নম্বর।

21নং চিত্রে 17নং টেবলের Histogram-টি অংকিত হইয়াছে। 'Best fit' Normal curve অংকিত করিতে হইলে প্রথমে Y_o বাহির করিতে হইবে। Normal curve-এর সূত্রে আছে $Y = \frac{N}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$.

Y_o, Mean-এর উপর অংকিত Y অক্ষের উচ্চতা। সেখানে x = o ; সুতরাং $Y_o = \frac{N}{\sigma\sqrt{2\pi}}$ (সূত্রে σকে ধাপ-এককে রাখিতে হয়। কারণ × অক্ষের এককও ধাপ-এককে থাকে)। 17নং টেবলে ঐ distribution-এর $Y_o = \frac{303}{2 \cdot 1 \times 2 \cdot 51} = 57 \cdot 5$. Y_o—এর মান বাহির করার পর আমরা Mean



21নং চিত্র : 17নং টেবলের Histogram এবং তাহার উপর "best fit" Normal curve.

হইতে σ—দূরত্বে Y—অক্ষের উচ্চতা বাহির করিতে পারি। B টেবল হইতে আমরা পাই যে,

Mean হইতে ± ·5σ দূরে Y অক্ষের উচ্চতা $= \cdot 88250 \times 57 \cdot 5 = 50 \cdot 60$
" " ± 1·0σ " " " " $= \cdot 60653 \times 57 \cdot 5 = 35 \cdot 02$
" " ± 1·5σ " " " " $= \cdot 32465 \times 57 \cdot 5 = 18 \cdot 77$
" " ± 2·0σ " " " " $= \cdot 13534 \times 57 \cdot 5 = 7 \cdot 76$
" " ± 2·5σ " " " " $= \cdot 04394 \times 57 \cdot 5 = 2 \cdot 53$
" " ± 3·0σ " " " " $= \cdot 01111 \times 57 \cdot 5 = \cdot 64$

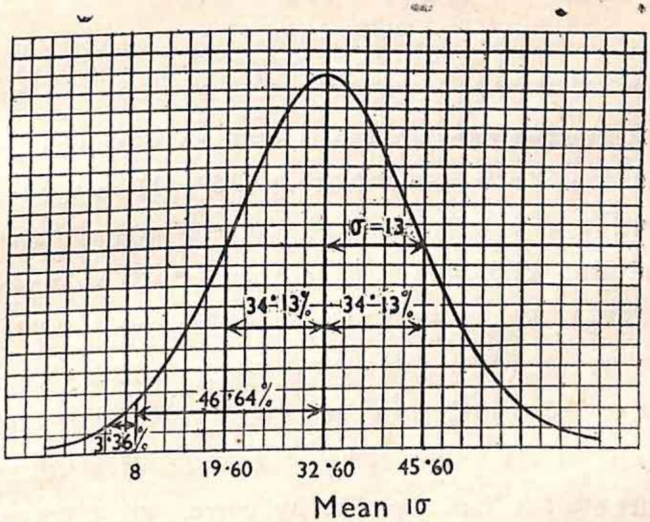
এখন Mean হইতে ±5σ দূরে অর্থাৎ ±10·48 একক দূরে অর্থাৎ × অক্ষে 91·11 এবং 70·15 এককের উপর 50·66 Y-উচ্চতা উঠিবে। এইরূপে যথাক্রমে M±σ অর্থাৎ 101·58 এবং 59·68 এককের উপর 35·02

Y-উচ্চতা উঠিবে ইত্যাদি। ঐ উচ্চতার শীর্ষবিন্দুগুলি পর্যায়-ক্রমে যোগ করিলে যে curve হইবে—তাহাই ঐ Histogram-এর 'Best fit' Normal curve.

(D) Normal probability curve-এর প্রয়োগ

এই অংশে আমরা কতকগুলি সমস্যার মাধ্যমে Normal curve-এর প্রয়োগ সম্বন্ধে আলোচনা করিব। ইহা কিন্তু ধরিয়া লওয়া হয় যে distribution-টি Normal অথবা Normal-এর কাছাকাছি। এই সমস্যাগুলির সমাধানে Table A খুবই সহায়তা করিবে।

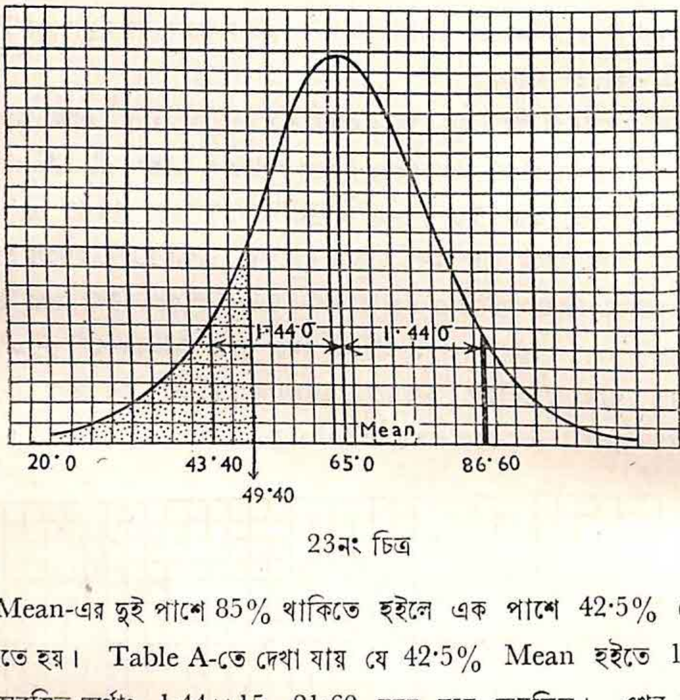
(ক) প্রথমতঃ প্রদত্ত দুই সীমার মধ্যে শতকরা কতগুলি লোক আছে—Normal curve তাহার হিসাব দেয়। মনে করা গেল—কোন distribution-এ Mean 32·60, $\sigma = 13{\cdot}0$; তাহা হইলে 19·60 এবং 45·60 এর মধ্যে কতগুলি লোক আছে তাহা নির্ণয় করিতে হইবে। এখন 19·60 নম্বরটি Mean 32·60 হইতে 13·0 নম্বর নীচে এবং 45·60 Mean হইতে 13·0 নম্বর উপরে অবস্থিত। ∴ 19·60 এবং 45·60 নম্বর দুইটি Mean-এর $\pm\,\sigma$ দূরে অবস্থিত থাকায় উহাদের মধ্যস্থলে distribution-এর 68·26% লোক বিদ্যমান আছে। এই রূপে প্রদত্ত যে কোন দুই সীমার মাঝে শতকরা কত



22নং চিত্র

জন বিদ্যমান Table A দ্বারা তাহা নির্ণয় করা যায়। 8-এর নীচে শতকরা 3·36% জন আছে—তাহা 22নং চিত্রে দেখান হইয়াছে।

(খ) Normal curve প্রদত্ত শতকরা লোকসংখ্যার মাধ্যমে কোন নম্বরের সীমা বাহির করিতে সাহায্য করে। মনে করি কোন distribution-এর $M = 65{\cdot}00$ এবং $\sigma = 15{\cdot}00$. এখন মাঝামাঝি 85% লোক থাকিতে পারে এমন নম্বরের দুইটি সীমা বাহির করিতে হইবে।

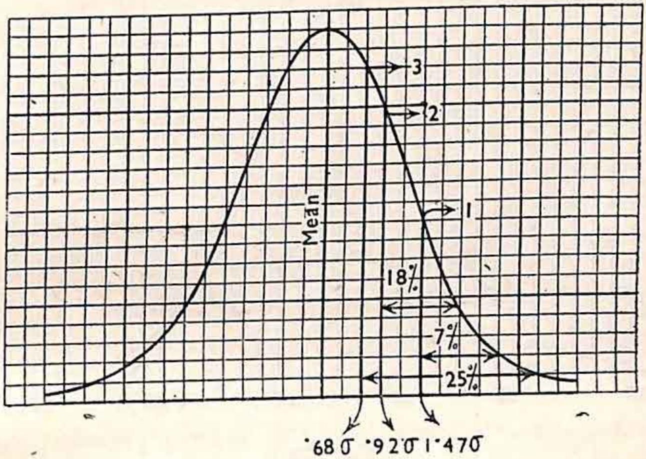


23নং চিত্র

Mean-এর দুই পাশে 85% থাকিতে হইলে এক পাশে 42·5% লোক থাকিতে হয়। Table A-তে দেখা যায় যে 42·5% Mean হইতে $1{\cdot}44\sigma$ দূরে অবস্থিত অর্থাৎ $1{\cdot}44 \times 15 = 21{\cdot}60$ নম্বর দূরে অবস্থিত। এখন সীমা হইবে Mean $\pm 21{\cdot}60$ অর্থাৎ 43·40 হইতে 86·60-এর মধ্যে ঐ distribution-এর মাঝখানের 85% লোক অবস্থিত। সর্বনিম্ন 15% লোক সংখ্যার সীমা কত তাহাও চিত্রে (23নং চিত্রে) দেখান হইয়াছে। 20·0 হইতে 49·40 সীমার মধ্যে সর্বনিম্ন 15% লোক বিদ্যমান।

(গ) অভীক্ষার প্রশ্ন বা সমস্যা নির্বাচন করিতে হইলে কাঠিন্যের ক্রম অনুসারে সাজাইতে Normal probability curve. সাহায্য করে মনে করি 1নং প্রশ্নে শতকরা 7%, 2নং প্রশ্নে 18% এবং 3নং প্রশ্নে 25% পরীক্ষার্থী ঠিক উত্তর করিয়াছে। এখন এই তিনটি প্রশ্নের কাঠিন্য কত তাহা নির্ণয় করিতে হইবে।

24নং চিত্রে Mean-এর ডান দিকের অংশের প্রান্ত হইতে 1নং, 2নং, এবং 3নং প্রশ্নের এলাকা সাজাইতে হইবে। Table A-এর সাহায্যে দেখা



24নং চিত্র

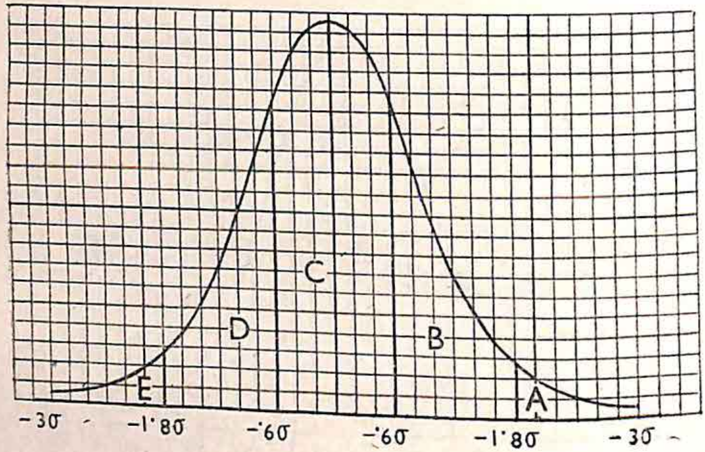
যায় (50% – 7%) বা 43% এলাকা Mean হইতে 1·47 σ দূরে অবস্থিত। এইরূপে দ্বিতীয়টি Mean হইতে ·92σ দূরে এবং তৃতীয়টি Mean হইতে ·68σ দূরে অবস্থিত। এখন ছক আকারে লিখিলে প্রশ্ন তিনটির আপেক্ষিক কাঠিন্য বুঝা যাইবে।

প্রশ্ন	উত্তীর্ণ	σ মান	তফাৎ
1	7%	1·47σ	—
2	18%	·92σ	·55σ
3	28%	·68σ	·24σ

(ঘ) আবার কোন গুণ যদি Normally distributed থাকে তাহা হইলে কোন দলকে নিজেদের ক্ষমতানুযায়ী কয়েকটি ক্ষুদ্রদলে বিভক্ত করিতে পারা যায়।

মনে করি কোন শ্রেণীতে মোট 150 জন ছাত্র পড়ে। ইহাদিগকে ক্ষমতানুযায়ী 5টি দলে ভাগ করিতে হইবে। মনে করি যে গুণের জন্য তাহাদিগকে ভাগ করিতে হইতেছে তাহা Normally distributed. গোটা Normal curve-এর X-অক্ষে 6σ দৈর্ঘ্য বিদ্যমান। সুতরাং পাঁচ ভাগ করিলে প্রতি

ভাগে 1·2σ করিয়া পড়িবে। মাঝখানের 1·2σ টিকে Mean-এর দুই পাশে ·6σ করিয়া লেখা হইল। অন্যান্য ভাগ চিত্রানুযায়ী করা হইল। ডান দিক



25নং চিত্র

হইতে ভাগগুলির নাম দেওয়া হইল—A, B, C, D এবং E : এখন Table A হইতে প্রতি এলাকায় শতকরা কত লোক আছে তাহা নির্ণয় করা হইল এবং নীচে তাহার হিসাব দেওয়া হইল।

	দল				
	A	B	C	D	E
প্রতি দলের শতকরা হিসাব	3·5	23·8	45	23·8	3·5
প্রতি দলের ছাত্র সংখ্যা (150 এর মধ্যে)	6 বা 5	36	67	36	5 বা 6

(E) **Normal curve-এর বৈশিষ্ট্য**

উপরে Normal curve সম্বন্ধে যে সকল আলোচনা হইয়াছে তাহা হইতে সংক্ষেপে curve-এর বৈশিষ্ট্যগুলি এখানে দেওয়া হইল :—

(ক) Normal curve-এর $Y = \frac{N}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$ সমীকরণ হইতে স্পষ্ট দেখা যায় যে ইহা একটি Asymtotic curve. X-অক্ষ অনন্তে curve-টির দুই প্রান্তে স্পর্শক। নতুবা কখনও curve-টি X-অক্ষকে ছেদ করে না।

(খ) x=0 হইলে, Y এর উচ্চতা সর্বাপেক্ষা বড় হয়। তাই Mean-এর

উপর Y-এর উচ্চতা বৃহত্তম। $x=\infty$ হইলে Y-এর মান ন্যূনতম হয়—এবং সেটি সম্ভব হয় অনন্তে।

(গ) Normal curve-এ M, Mdn, এবং Mo একই বিন্দুতে অবস্থিত।

(ঘ) Normal curve Mean-এর দুই পার্শ্বে সদৃশ রেখাদ্বারা সীমাবদ্ধ।

(ঙ) Normal curve-টি বামদিক হইতে প্রথমে convex, পরে concave হইয়া plateau-তে পৌছায়। পরে নিম্নগামী হইয়া প্রথমে concave এবং পরে convex হইয়া X-অক্ষে অনন্তে গিয়া মিশে।

(চ) Mean-এর বামদিকে যে বিন্দুতে curve-টি convex হইতে concave হইতেছে এইং দক্ষিণদিকে যে বিন্দুতে curve-টি concave হইতে convex হইতেছে—সেই বিন্দুদ্বয়ের প্রত্যেককে point of inflection বলে। ঐ দুই বিন্দু হইতে X-অক্ষের উপর যদি দুই পাশে দুইটি লম্ব অংকিত করা যায় তবে তাহা Mean হইতে $\pm\sigma$ দূরত্বে ছেদ করে।

(ছ) Normal curve এ $M\pm\sigma$-এর মধ্যে কোন distribution এর মাঝের 68·26%, $M\pm2\sigma$-এর মধ্যে মাঝের প্রায় 95% এবং $M\pm3\sigma$-এর মধ্যে মাঝের প্রায় 99·73% লোক থাকে।

(জ) Normal curve হইতে Table A দ্বারা যে কোন নম্বরের সীমা হইতে শতকরা কত লোক আছে, তাহা বলা যায়। বিপরীতক্রমে কোন শতকরা সীমাতে কত নম্বর আছে তাহাও বলা যায়।

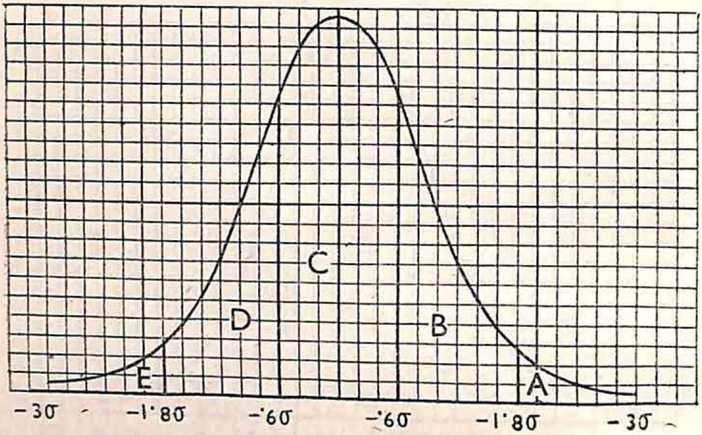
(ঝ) Normal curve-এর Skewness (22নং সূত্রে 0 এবং Kurtosis (24নং সূত্রে) ·263.

(ঞ) Normal curve, probability বা chance-এর curve. Prabability Polygon curve চরম সময়ে (যখন $N=\infty$) Normal curve পাওয়া যায়। ইহা আদর্শ লেখচিত্র। ইহার দ্বারা পরিসংখ্যানে বিভিন্ন বিজ্ঞানকে সৃষ্টিমূলক ভাবে প্রয়োগ করা সম্ভব হইয়াছে।

(F) Non-Normality-র কারণ ঃ

পূর্বেই আলোচনা করা হইয়াছে যে, distribution যে সব সময় Normal বা তাহার কাছাকাছি হইবে তাহার কোন নিশ্চয়তা নাই। তবে সত্যই যদি Skewness বা Kurtosis এত বেশী হইয়া পড়ে যে chance error দিয়াও ব্যাখ্যা করা যায় না—তখন SK বা Ku এর কারণ উদ্ঘাটন করা প্রয়োজন। হিসাবে অগণিত কারণ থাকিলেও গবেষকের কাছে কয়েকটি বিশেষ কারণ

ভাগে $1{\cdot}2\sigma$ করিয়া পড়িবে। মাঝখানের $1{\cdot}2\sigma$ টিকে Mean-এর দুই পাশে $\cdot6\sigma$ করিয়া লেখা হইল। অন্যান্য ভাগ চিত্রানুযায়ী করা হইল। ডান দিক



25নং চিত্র

হইতে ভাগগুলির নাম দেওয়া হইল—A, B, C, D এবং E : এখন Table A হইতে প্রতি এলাকায় শতকরা কত লোক আছে তাহা নির্ণয় করা হইল এবং নীচে তাহার হিসাব দেওয়া হইল।

	দল				
	A	B	C	D	E
প্রতি দলের শতকরা হিসাব	3·5	23·8	45	23·8	3·5
প্রতি দলের ছাত্র সংখ্যা (150 এর মধ্যে)	6 বা 5	36	67	36	5 বা 6

(E) Normal curve-এর বৈশিষ্ট্য

উপরে Normal curve সম্বন্ধে যে সকল আলোচনা হইয়াছে তাহা হইতে সংক্ষেপে curve-এর বৈশিষ্ট্যগুলি এখানে দেওয়া হইল :—

(ক) Normal curve-এর $Y = \frac{N}{\sigma\sqrt{2\pi}} . e^{-\frac{x^2}{2\sigma^2}}$ সমীকরণ হইতে স্পষ্ট দেখা যায় যে ইহা একটি Asymtotic curve. X-অক্ষ অনন্তে curve-টির দুই প্রান্তে স্পর্শক। নতুবা কখনও curve-টি X-অক্ষকে ছেদ করে না।

(খ) x=0 হইলে, Y এর উচ্চতা সর্বাপেক্ষা বড় হয়। তাই Mean-এর

উপর Y-এর উচ্চতা বৃহত্তম। $x = \infty$ হইলে Y-এর মান ন্যূনতম হয়—এবং সেটি সম্ভব হয় অনন্তে।

(গ) Normal curve-এ M, Mdn, এবং Mo একই বিন্দুতে অবস্থিত।

(ঘ) Normal curve Mean-এর দুই পার্শ্বে সদৃশ রেখাদ্বারা সীমাবদ্ধ।

(ঙ) Normal curve-টি বামদিক হইতে প্রথমে convex, পরে concave হইয়া plateau-তে পৌছায়। পরে নিম্নগামী হইয়া প্রথমে concave এবং পরে convex হইয়া X-অক্ষে অনন্তে গিয়া মিশে।

(চ) Mean-এর বামদিকে যে বিন্দুতে curve-টি convex হইতে concave হইতেছে এইং দক্ষিণদিকে যে বিন্দুতে curve-টি concave হইতে convex হইতেছে—সেই বিন্দুদ্বয়ের প্রত্যেককে point of inflection বলে। ঐ দুই বিন্দু হইতে X-অক্ষের উপর যদি দুই পাশে দুইটি লম্ব অংকিত করা যায় তবে তাহা Mean হইতে $\pm\sigma$ দূরত্বে ছেদ করে।

(ছ) Normal curve এ $M \pm \sigma$-এর মধ্যে কোন distribution এর মাঝের 68·26%, $M \pm 2\sigma$-এর মধ্যে মাঝের প্রায় 95% এবং $M \pm 3\sigma$-এর মধ্যে মাঝের প্রায় 99·73% লোক থাকে।

(জ) Normal curve হইতে Table A দ্বারা যে কোন নম্বরের সীমা হইতে শতকরা কত লোক আছে, তাহা বলা যায়। বিপরীতক্রমে কোন শতকরা সীমাতে কত নম্বর আছে তাহাও বলা যায়।

(ঝ) Normal curve-এর Skewness (22নং সূত্রে 0 এবং Kurtosis (24নং সূত্রে) ·263.

(ঞ) Normal curve, probability বা chance-এর curve. Prabability Polygon curve চরম সময়ে (যখন $N = \infty$) Normal curve পাওয়া যায়। ইহা আদর্শ লেখচিত্র। ইহার দ্বারা পরিসংখ্যানে বিভিন্ন বিজ্ঞানকে সৃষ্টিমূলক ভাবে প্রয়োগ করা সম্ভব হইয়াছে।

(F) Non-Normality-র কারণ :

পূর্বেই আলোচনা করা হইয়াছে যে, distribution যে সব সময় Normal বা তাহার কাছাকাছি হইবে তাহার কোন নিশ্চয়তা নাই। তবে সত্যই যদি Skewness বা Kurtosis এত বেশী হইয়া পড়ে যে chance error দিয়াও ব্যাখ্যা করা যায় না—তখন SK বা Ku এর কারণ উদ্ঘাটন করা প্রয়োজন। হিসাবে অগণিত কারণ থাকিলেও গবেষকের কাছে কয়েকটি বিশেষ কারণ

বিশ্লেষণে ধরা পড়ে। এখানে Nou-Normality-র তেমনই বিশেষ কয়েকটি কারণ আলোচিত হইয়াছে।

(ক) প্রথম কারণটি হইল বাছাই। বাছাই-এর মধ্যে ফাঁক থাকিলে curve-এর SK বা Ku থাকিবেই। খুব ভাল বা খুব খারাপ পরীক্ষার্থীর উপর অভীক্ষা প্রয়োগ করিলে Normal curve আশা করা যায় না—যতই ভালভাবে হিসাব মত অভীক্ষা প্রস্তুত হউক না কেন। বিভিন্ন ভাবাভাষী, বিভিন্ন সামাজিক ও অর্থনৈতিক পরিবেশে মানুষ, বেশী বয়সে নীচু শ্রেণীতে পড়া বা হঠাৎ বৃদ্ধি পাওয়া ছেলেদের জন্য curve-টি Non-Normal হয়। Homogeneous দলের ফল Leptokurtic এবং Heterogeneous দলের ফল স্বাভাবিক ভাবে Platykurtic হইবে।

(খ) যে গুণ পরিমাপ করিতে হইবে তাহা যদি জনসাধারণে Normally বিস্তৃত না থাকে তবে কোন ক্রমেই Normal curve পাওয়া সম্ভব নয়। যদি কোন মুদ্রার H এবং T পড়ার সম্ভাবনা প্রত্যেকটিতে $\frac{1}{2}$ না থাকে—যদি $H = \frac{2}{3}$ এবং $T = \frac{1}{3}$ হয় তাহা হইলে Normal curve পাওয়া যাইবে না। তখন $(H+T)^3 = H^3 + 3H^2T + 3HT^2 + T^3$ সূত্রে $H = \frac{2}{3}$ এবং $T = \frac{1}{3}$ স্থাপন করিলে—

(i) $H^3 = (\frac{2}{3})^3 = \frac{8}{27}$

$3H^2T = 3(\frac{2}{3})^2 \frac{1}{3} = \frac{12}{27}$

$3HT^2 = 3(\frac{2}{3})(\frac{1}{3})^2 = \frac{6}{27}$

$T^3 = (\frac{1}{3})^3 = \frac{1}{27}$

(ii) ইহাদিগকে frequency distribution-এ সাজাইলে

H এর সংখ্যা	f
3	8
2	12
1	6
0	1

উপরে স্পষ্টই দেখা যাইতেছে যে উক্ত সর্ত্তে ($H = \frac{2}{3}$ এবং $T = \frac{1}{3}$) লেখচিত্রটি কখনও Normal হইতে পারে না।

(গ) Non-Normality-র আর একটি কারণ হইল—কঠিন বা সহজ প্রশ্নের অবতারণা। প্রশ্ন যদি ক্ষমতার অনুপাতে খুব সহজ হয় তবে স্কেলের শেষ প্রান্ত ভারী হইবে—অর্থাৎ অধিক সংখ্যক ছাত্র বেশী নম্বর পাইবে। আবার প্রশ্ন খুব কঠিন হইলে স্কেলের নিম্নপ্রান্ত ভারী হইবে—অর্থাৎ অধিক সংখ্যক

ছাত্র কম নম্বর পাইবে। সুতরাং বিশেষ চিন্তা করিয়া সংশুদ্ধ অভীক্ষা পরীক্ষার্থীর উপযোগী করিলে তবেই Normal curve পাওয়া যাইবে।

(ঘ) ইহা ছাড়া অভীক্ষা প্রস্তুত এবং অভীক্ষার মান প্রদানের তারতম্যের জন্যও লেখচিত্র Normal হয় না। অভীক্ষা সময়, প্রয়োগকালে উপদেশ প্রদান, উৎসাহ দানের তারতম্যের জন্যও Non-Normality দেখা যায়। যে ছাত্র খুব কম নম্বরের অধিকারী উৎসাহ প্রদানের জন্য সে বেশীও পাইতে পারে। আবার ইহার বিপরীত ফলও দৃষ্ট হয়।

Exercise VI

1. What percentage of a normal distribution is included between the

(a) Mean and $1 \cdot 72\ \sigma$ (d) $-1 \cdot 7$ PE and $1 \cdot 5$ PE
(b) Mean and $-\cdot 79\ \sigma$ (e) $\cdot 56\ \sigma$ and $1 \cdot 93\ \sigma$
(c) $-\cdot 64\ \sigma$ and $\cdot 46\ \sigma$ (f) $-1 \cdot 6$ PE and $-2 \cdot 7$ PE.

2. In a normal distribution

(a) Determine P_{21}, P_{42}, P_{61} and P_{91} in σ units.

(b) What are the percentile ranks of scores at $-1 \cdot 02\sigma$, $-\cdot 34\ \sigma$, $+1 \cdot 64\ \sigma$?

3. (a) Compute measures of skewness and of Kurtosis for the first two frequency distribution in Exercise—V, problem 3.

(b) Fit normal probability curves to these same distributions.

(c) For each distribution, compare the percentage of cases lying between $\pm 1\sigma$ with the 68·28% found in the normal distribution.

4. Suppose that the height of maximum ordinate (Y_0) in a normal curve is 50. What is the height to the nearest integer of the ordinate at the x/σ point which cuts off the top 13% of the distribution ? top 43% ? bottom 8% ? (use table A and B)

5. In a sample of 1000 cases the mean of the certain test is 15·20 and σ is 3·28.

Assuming normality of distribution

(a) How many individuals score between 11 and 18 ?

(b) How many score above 19 ? below 5 ?

(c) What are the chances that any individual selected at random will score above 15 ?

সপ্তম অধ্যায়

"রৈখিক অনুবন্ধ"

পূর্বের অধ্যায় সমূহে কোন দলের প্রতিভূ, বিস্তৃতি, লেখচিত্র অঙ্কন প্রভৃতি তথ্য সংগৃহীত হইয়াছে, কিন্তু একটি দলের নম্বরের সংগে অন্য দলের নম্বরের সাদৃশ্য বা বৈসাদৃশ্য প্রভৃতি আলোচনা করা দরকার। একটি দলের বিভিন্ন গুণের মধ্যে কি রকম সম্বন্ধ বিদ্যমান বা একটি গুণ দ্বারা বিচার করিলে বিভিন্ন দলের মধ্যে কোন সম্পর্ক আছে কিনা তাহা জানা দরকার। একটি ছাত্র বুদ্ধির পরীক্ষায় ভাল করিল—সে স্মৃতির পরীক্ষায় ভাল করিবে কি না সে ভবিষ্যৎবাণী করা সম্ভব হইবে—যদি বুদ্ধি ও স্মৃতির মধ্যে কি সম্পর্ক বর্তমান তাহা পূর্বে জানিতে পারি। অনুবন্ধ পদ্ধতির মাধ্যমে ঐ বিষয়ে সম্যক্ জ্ঞান লাভ সম্ভবপর।

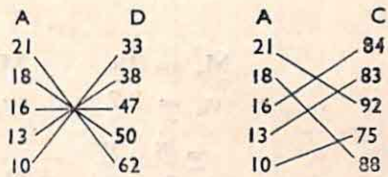
দুই বা ততোধিক গুণের মধ্যে যে সম্পর্ক সাদৃশ্য বা বৈসাদৃশ্য—(Correspondence) তাহারই নাম অনুবন্ধ। কতটা সম্পর্ক তাহা সংখ্যা দ্বারা প্রকাশ করা হয়। যত ডিগ্রী সম্পর্ক তত ডিগ্রীকে অনুবন্ধ সহগ বলে।

অনুবন্ধ সহগ ধনাত্মক, ঋণাত্মক বা শূন্য হইতে পারে। সর্বাপেক্ষা বৃহত্তম সহগ হইতেছে 1·00. ক্ষুদ্রতম −1·00. সুতরাং বুঝা যাইতেছে যে অনুবন্ধ-সহগ −1·00 হইতে শূন্যকে মাঝে রাখিয়া +1·00 সীমার মধ্যে অবস্থিত। নীচে একটি উদাহরণ দেওয়া হইল। কোন বৃত্তের ক্ষেত্রফল πr^2। এখন যে কোন ছোট বা বড় বৃত্ত অংকিত করিয়া তাহার ক্ষেত্রফলকে r^2 দ্বারা ভাগ করিলে প্রতিবারই π বা প্রায় 3·14 হইবে। সুতরাং বৃত্তের ক্ষেত্রফলের সহিত r^2 এর সম্পর্ক খাঁটি এবং অনুবন্ধ-সহগ 1·00। আবার কোন কাজে লোক লাগান হইল। যদি বেশী লোক কাজ করে তবে কাজটি শেষ হইতে কম দিন লাগিবে এবং বিপরীতক্রমে কম লোক লাগিলে বেশী দিনে কাজটি সম্পূর্ণ হইবে। এখানে লোকসংখ্যার সহিত দিন সংখ্যার একটি খাঁটি সম্পর্ক বিদ্যমান। এখানে অনুবন্ধ সহগ −1·00· এ ছাড়া ঐ দুই সম্পর্ক যখন দেখা যায় না তখন সহগ হইবে 0·00.

পদমর্যাদার মাধ্যমে সহগের বিভিন্ন অবস্থা নীচে দেখান হইল। পাঁচজন ছাত্রের চারিটি বিষয়ের পরীক্ষার ফল A, B, C, D সারিতে দেওয়া হইয়াছে। প্রথম তুলনায় A সারিতে মানের ক্রমানুসারে নম্বর সাজান হইয়াছে এবং তাহাদের বরাবর B সারিতে তাহাদের B বিষয়ের ফল লেখা হইয়াছে। এইরূপে দ্বিতীয় তুলনা এবং তৃতীয় তুলনা করা হইয়াছে।

প্রথম দ্বিতীয় তৃতীয়

	A	B
a	21—	85
b	18—	66
c	16—	50
d	13—	40
e	10—	32



r=1·00 r=—1·00 r=শূন্যের কাছাকাছি

এখন প্রথম তুলনায় দেখা যাইতেছে যে A সারিতে যে প্রথম হইয়াছে B সারিতেও সে প্রথম হইয়াছে; এবং B সারির প্রত্যেক নম্বরটি A সারির ন্যায় পর পর আসন লইয়া বসিয়া অছে। অর্থাৎ A সারির সর্ব নিম্ন নম্বরের অধিকারী B সারিতেও সর্বনিম্ন নম্বর পাইয়াছে। তাই সম্বন্ধটি সোজা রেখা দ্বারা প্রকাশ করা হইয়াছে। সেইজন্য সেখানকার অনুবন্ধ সহগ বা r খাঁটি এবং 1·00.

আবার দ্বিতীয় তুলনাটি প্রথমটির ঠিক বিপরীত। এখানে A সারিতে যে সর্বোচ্চ নম্বর পাইয়াছে B সারিতে সেই সর্বনিম্ন নম্বর পাইয়াছে। এইরূপে সরলরেখাগুলি একটি বিন্দুতে ছেদ করিয়াছে। সেইজন্য এখানে সম্বন্ধও বিপরীত। r এখানে খাটি এবং —1·00।

তৃতীয় তুলনাটিতে কোন সম্পর্ক নাই। কারণ '+' এবং '—' সম্বন্ধ কাটাকুটি হইয়া ইহা সম্পর্কশূন্য হইয়াছে। তাই এখানে r=0·00।

A. অনুবন্ধ-সহগ ঃ

যতগুলি পদ্ধতির মাধ্যমে অনুবন্ধ-সহগ বা r বাহির করা হয় তন্মধ্যে Product Moment পদ্ধতি একটি। Product Moment হইতেছে এমন একটি অনুপাত—যাহা কোন এক variable এর পরিবর্তন আনয়ন করে বা পরিবর্তনের উপর নির্ভর করে। নীচের একটি উদাহরণের দ্বারা বিষয়টি পরিষ্কার হইবে।

1	2	3	4	5	6	7	8	9
ছাত্র	অংকের নম্বর (X)	ইংরাজীর নম্বর (Y)	x	y	xy	$\frac{x}{\sigma_x}$	$\frac{y}{\sigma_y}$	$\frac{x\,y}{\sigma_x\sigma_y}$
A	17	36	4	5	20	1·1	·48	·53
B	13	46	0	15	0	·00	1·43	·00
C	15	35	2	4	8	·55	·38	·21
D	14	26	1	− 5	−5	·27	−·48	−·13
E	6	12	−7	−19	133	−1·92	−1·81	3·48
								4·09

$$M_x = 13 \qquad M_y = 31$$
$$\sigma_x = 3{\cdot}7 \qquad \sigma_y = 10{\cdot}5$$

$$\text{অনুবন্ধ সহগ} = \frac{\Sigma \frac{x}{\sigma_x} \cdot \frac{y}{\sigma_y}}{N} = \frac{4{\cdot}09}{5} = {\cdot}82$$

18 নং টেবল : পাঁচটি ছাত্রের অংকের ও ইংরাজীর নম্বরের অনুবন্ধ সহগ।

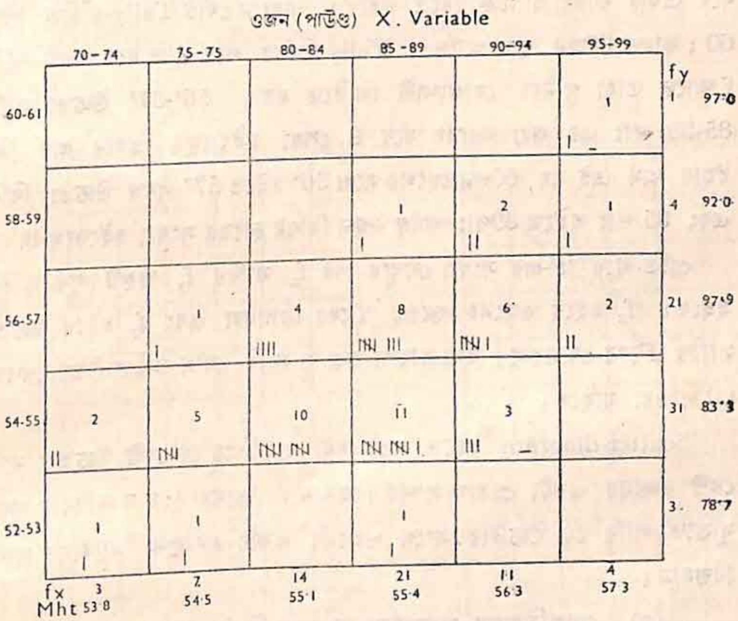
18নং টেবিল হইতে স্পষ্ট দেখা যাইতেছে যে অংকের নম্বরের সংগে ইংরাজীর নম্বরের প্রত্যক্ষ সম্বন্ধ বিদ্যমান। কেবল দুই একটি ক্ষেত্রে ব্যতিক্রম আছে এখন মোট r কত তাহা নির্ণয় করিতে হইবে। X এবং Y স্তম্ভের পরে x এবং y 4 এবং 5 স্তম্ভে রাখা হইয়াছে। প্রত্যেকটি সংখ্যা ঐ দলের Mean হইতে কত দূরে তাহাই ঐ স্তম্ভদ্বয়ে দেখান হইয়াছে। পরবর্তী স্তম্ভে (6) x, y এর গুণফল, 7 এবং 8 নং স্তম্ভে নম্বরগুলির Mean হইতে দূরত্বকে ঐ দলের σ-এককে লইয়া যাওয়া হইয়াছে। 9নং স্তম্ভে 7 এবং 8নং স্তম্ভের গুণফল দেওয়া হইয়াছে। 9নং স্তম্ভের যোগফলকে N দ্বারা ভাগ করিলে r পাওয়া যাইবে।

Mean হইতে দূরত্বের (x বা x এর অন্য কোন ঘাত) যোগফলকে N দ্বারা ভাগ করিলে 'Moment' পাওয়া যায়। এখন x এবং y দূরত্বগুলিকে নিজ নিজ σ—এককে লইয়া গুণ করিয়া গুণফলকে N দ্বারা ভাগ করিলে Product moment r পাওয়া যায়। তাই $r = \Sigma \frac{\frac{x}{\sigma_x} \cdot \frac{y}{\sigma_y}}{N}$ ....(25)

আবিষ্কারক Karl Pearson-এর নাম অনুসারে এই r, Pearsonian coefficient নামে খ্যাত।

B (ক) **Scatter diagram এবং অনুবন্ধ :** যখন N এর সংখ্যা বেশ

ছোট থাকে তখন অনুপাত পদ্ধতি বা Product moment পদ্ধতি বেশ কার্যকরী। কিন্তু N বেশী হইলে ঐ পদ্ধতিতে r বাহির করা সময় ও কষ্ট সাপেক্ষ। তাই N বেশী সংখ্যক হইলে



26নং চিত্র: Scatter diagram-এ 60 জন ছেলের উচ্চতা ও তাহাদের ওজনের Tally করা হইয়াছে।

26নং চিত্রের মত একটি তালিকা করিয়া তাহাতে শ্রেণীবিহীন দুইটি variable-কে সাজাইয়া শ্রেণীবদ্ধ করা হয়। উক্ত তালিকাকে Scatter diagram বা Scatter-gram বলে। দুইটি variable-উচ্চতা ও ওজন—একই তালিকাতে সাজান হইয়াছে বলিয়া ইহাকে bivariate distribution বলা হয়—আবার Scatter-gram-কে অনেকে অনুবন্ধ তালিকাও বলেন। 26নং চিত্রে বামদিকে নীচ হইতে উপর পর্যন্ত ছাত্রদের উচ্চতা ইঞ্চিতে কয়েকটি ধাপে এবং উপরে বাম হইতে দক্ষিণে তাহাদের ওজন পাউণ্ডে কয়েকটি ধাপে প্রকাশ করা হইয়াছে। মনে করি কোন ছাত্রের উচ্চতা 58·7" এবং তাহার ওজন 86 পাউণ্ড। তাহা হইলে উপর হইতে দ্বিতীয় সারি বরাবর 85-89 পাউণ্ডের স্তম্ভে একটি Tally চিহ্ন '।' দিতে হইবে। এইরূপ যদি অন্য

ছাত্রের উচ্চতা 54·8″ এবং ওজন 93 পাউণ্ড হয় তবে Scatter diagram-এ নীচ হইতে দ্বিতীয় সারি বরাবর 90-94 পাউণ্ডের স্তম্ভে একটি Tally চিহ্ন '।' দিতে হইবে। এইরূপে প্রতি ছাত্রের উচ্চতা এবং তাহার ওজন বরাবর ঘরে একটি করিয়া '।' চিহ্ন দিতে হইবে। এখানে মোট Tally-র চিহ্ন হইবে 60 ; কারণ 60জন ছাত্র আছে। Tally চিহ্নের পরে প্রতি ঘরে কয়টি করিয়া f আছে তাহা গনিয়া যোগফলটি লিখিতে হয়। 56″-57″ উচ্চতা সারির 85-89 পাঃ এর স্তম্ভ বরাবর ঘরে 8 লেখা হইয়াছে। ইহার অর্থ কি? ইহার অর্থ এই যে 60জন ছেলের মধ্যে 56″ হইতে 57″ পর্যন্ত উচ্চতা বিশিষ্ট এবং 85 পাঃ হইতে 89পাঃ পর্যন্ত ওজন বিশিষ্ট ছাত্রের সংখ্যা হইতেছে 8.

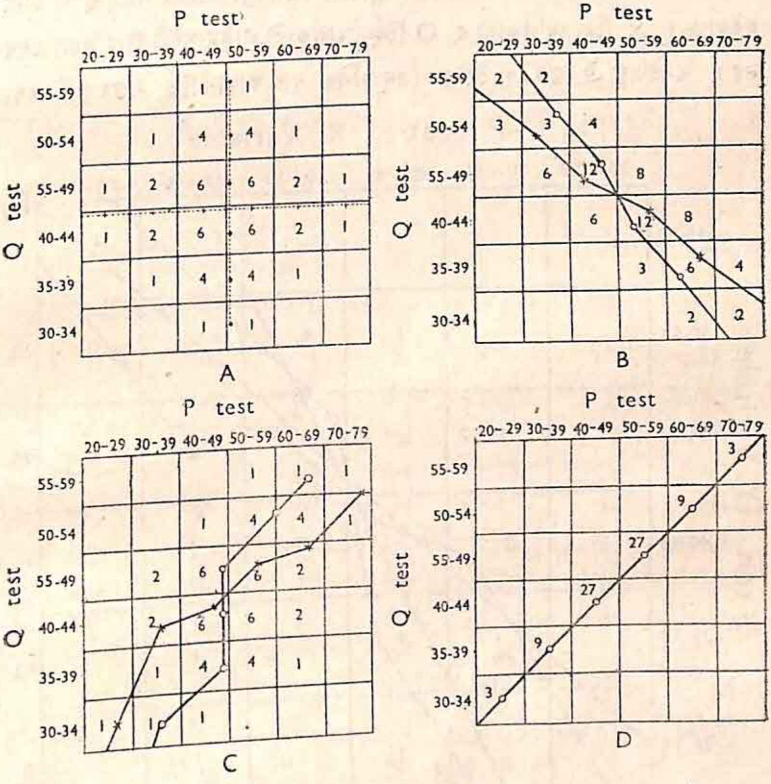
প্রতি ঘরে 'f'-এর সংখ্যা লেখার পর f_x সারির f_y স্তম্ভটি পূর্ণ করিতে হইবে। f_x হইবে ওজনের স্তম্ভের 'f'দের যোগফল এবং f_y হইবে উচ্চতার সারির 'f'দের যোগকল। বলা বাহুল্য স্তম্ভ বা সারি যোগ করিয়া উভয় ক্ষেত্রেই 60 পাওয়া যাইবে।

Scatter-diagram হইতে স্পষ্ট দেখা যাইতেছে যে বেশী উচ্চতার সংগে বেশী ওজনের একটা প্রত্যক্ষ সম্পর্ক বিদ্যমান। আমরা হিসাব করিবার পূর্বেই বুঝিতে পারি যে উচ্চতার সংগে ওজনের একটা ধনাত্মক অনুবন্ধের সম্পর্ক বিদ্যমান।

(খ) **লেখচিত্রের সাহায্যে অনুবন্ধ নির্ণয়ঃ** নীচে চারিটি Scatter-gram A,B,C,D দেওয়া হইল। ইহাদের সারিতে P-test এবং স্তম্ভে Q-test-এর ফল রাখা হইয়াছে। (27নং চিত্র)।

প্রতি P-স্তম্ভে Q এর প্রতি Array-এর Mean-এর স্থানকে '×' চিহ্নদ্বারা প্রকাশ করা হইয়াছে। এই '×' চিহ্নগুলি যোগ করিয়া একটি রেখা পাওয়া যায়। ইহার নাম Regression line. ঐ রেখা দ্বারা প্রদত্ত P-সীমানার মধ্যে Q-এর Mean কি ভাবে পরিবর্তিত হইতেছে তাহাই বুঝান হইতেছে। এইরূপে P-array এর Meanগুলির (প্রতি Q-সারির জন্য) স্থান 'O' চিহ্ন দ্বারা প্রকাশ করা হইয়াছে। ঐগুলি যোগ করিয়া আর একটি regression line পাওয়া যায়। শেষোক্ত রেখাদ্বারা প্রদত্ত Q সীমানার মধ্যে P-এর Mean কিভাবে পরিবর্তিত হইতেছে—তাহাই সূচিত হইয়াছে। এই দুইটি রেখার দ্বারা আমরা P এবং Q-এর মধ্যে সম্পর্ক কি রকম বা কতটুকু তাহা নির্ণয় করিতে সক্ষম হই।

কাঁচির পাতের মত অবস্থিত Regression line দুইটির অবস্থানের উপর দুই অভীক্ষার সম্বন্ধ নির্ভরশীল। রেখা দুইটি যত কাছাকাছি থাকিবে সম্বন্ধ ততই বেশী হইবে। যদি বামদিকের নীচ হইতে ডানদিকের উপর বরাবর অবস্থিত থাকে তবে r ধনাত্মক হইবে (চিত্র 27-C). এইরূপ অবস্থায় রেখা দুইটি উপর্যুপরি থাকিলে r = 1·00 হইবে (চিত্র 27-D)



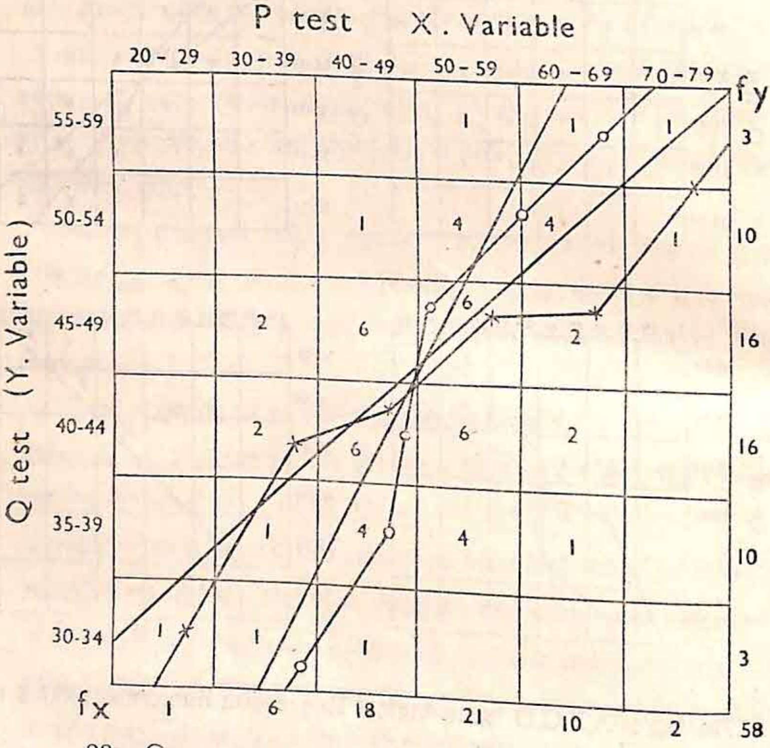
27নং চিত্র ঃ A,B,C,D চিত্রের সাহায্যে Regression line দেখান হইয়াছে।

ইহারই বিপরীত অবস্থা হইতেছে চিত্র 27-B—যেখানে ডানদিকের নীচ হইতে উপরের বাম কোণ বরাবর regression line দুইটি অবস্থিত। এখানে r ঋণাত্মক। এইরূপ অবস্থায় রেখা দুইটি মিশিলে r = —1·00 হইবে।

একটি অদ্ভুত চিত্রে (চিত্র-27-A) দুইটি variable P এবং Q-এর সম্বন্ধ দেখান হইয়াছে। এখানে লাইন দুইটি পরস্পরে লম্বভাবে অবস্থিত। সমগ্র

সীমানার মধ্যে Q-এর Mean-এর কোন পরিবর্তন নাই। একথা P-এর Mean-এর ক্ষেত্রেও প্রযোজ্য। এখানে r = ·00

সাধারণ ক্ষেত্রে চিত্র 27-A এবং D-এর ন্যায় regression line দুইটি সরলরেখাক্রমে অবস্থিত থাকে না। 28নং চিত্রে কি করিয়া regression লাইন দুইটিকে সরল রেখায় লইয়া যাওয়া হয় তাহা দেখান হইয়াছে। 28নং চিত্রে P-test এবং Q-test এর Scatter gram এ regression line দুইটি টানা হইয়াছে। × চিহ্ন যোগকারী বা O চিহ্ন যোগকারী লাইন দুইটি ঠিক সরল রেখ নয়। সেইজন্য ঐ দুই লাইনের বিন্দুগুলির যত কাছাকাছি সম্ভব দুই রেখা



28নং চিত্র : Regression line-কে সরল রেখায় পরিবর্তন।

যথাক্রমে টানা হইল। এখন পূর্বের লাইন দুইটি সরল রেখা নয় বলিয়া—নূতন অংকিত AB এবং CD সরল রেখাদ্বয়কে Regression line বলা হয়। Regression line-দ্বয়কে line-এর 'best fit' বলা হয়। ঐ line দ্বয় হইতে Scatter-gram-এ অবস্থিত বিন্দুগুলির দূরত্বের যোগফল সর্বাপেক্ষা কম।

ঐ নীতি অবলম্বন করিয়া লাইন দুইটি টানা হয়। পূর্বেই বলা হইয়াছে যে ঐ রেখা দুইটি (AB এবং CD) যতই কাছাকাছি অবস্থিত থাকিবে, P এবং Q অভীক্ষার মধ্যে r ততই বৃদ্ধি পাইবে।

C. Product Moment পদ্ধতির সাহায্যে r নির্ণয়।

(a) যখন নম্বরগুলি শ্রেণীবিহীনভাবে থাকে এবং deviation গুলি প্রকৃত Mean হইতে লওয়া হয়, তখন $r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$ সূত্রের সাহায্যে অনুবন্ধ সহগ বাহির করা হয়। কিন্তু আমরা জানি যে $\sigma_x = \sqrt{\frac{\Sigma x^2}{N}}$ এবং $\sigma_y = \sqrt{\frac{\Sigma y^2}{N}}$.

σ_x এবং σ_y —এর মান ঐ সূত্রে স্থাপন করিলে

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (26)$$

নীচে উক্ত সূত্রের সাহায্যে 19নং টেবলে r বাহির করা হইয়াছে।

ছাত্র	বিষয় (X)	বিষয় (Y)	x	y	x^2	y^2	xy
তাপস	85	65	6·5	5·6	42·25	31·36	36·40
সঞ্জয়	72	58	−6·5	−1·4	42·25	1·96	9·10
পথিক	75	55	−3·5	−4·4	12·25	19·36	15·40
কুমার	82	59	3·5	−·4	12·25	·16	−1·40
সৌমেন	75	63	−3·5	3·6	12·25	12·96	−12·60
সুদীপ্ত	80	56	1·5	−3·4	2·25	11·56	−5·10
সুব্রত	83	62	4·5	2·6	20·25	6·76	11·70
জয়ন্ত	76	57	−2·5	−2·4	6·25	5·76	6·00
					150·00	89·88	59·50
					(Σx^2)	(Σy^2)	(Σxy)

$$M_x = 78{\cdot}5 \qquad M_y = 59{\cdot}4$$

$$\sigma_x = 4{\cdot}3 \qquad \sigma_y = 3{\cdot}4$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{59{\cdot}50}{\sqrt{150{\cdot}00 \times 89{\cdot}88}} = \frac{59{\cdot}50}{\sqrt{13482}}$$

$$= \frac{59{\cdot}50}{116{\cdot}1} = {\cdot}51$$

19নং টেবল : 8 জন ছাত্রলব্ধ শ্রেণীবিহীন নম্বরের সাহায্যে দুই বিষয়ের অনুবন্ধ সহগ নির্ণয় (দীর্ঘ পদ্ধতি)।

26নং সূত্রের সাহায্যে শ্রেণাবিহীন নম্বরের অনুবন্ধ সহগ বাহির করার নিয়ম নীচে প্রদত্ত হইল।

(ক) প্রথমে X এবং Y অভীক্ষায় প্রাপ্ত নম্বরের প্রকৃত Mean বাহির করা হইল। M_x এবং M_y-এর ফল যথাক্রমে 78·5 এবং 59·4 হইল।

(খ) এখন M_x হইতে X সারির প্রতিটি নম্বরের প্রকৃত দূরত্ব বাহির করিয়া x স্তম্ভে লেখা হইল। y স্তম্ভটি অনুরূপ ভাবে পূরণ করা হইল।

(গ) x এবং y স্তম্ভের প্রতিটি দূরত্বে বর্গ নির্ণয় করিয়া যথাক্রমে x^2 এবং y^2 স্তম্ভদ্বয়ে লেখা হইল এবং Σx^2 ও Σy^2 বাহির করা হইল।

(ঘ) একই সারিতে পাশাপাশি x এবং y এর গুণফল নির্ণয় করিয়া xy স্তম্ভটি পূর্ণ করা হইল এবং Σxy বাহির করা হইল।

(ঙ) এখন $\Sigma x^2 = 150{\cdot}00$, $\Sigma y^2 = 89{\cdot}88$ এবং $\Sigma xy = 59{\cdot}50$-এর মান 26নং সূত্রে স্থাপন করিয়া r-এর মান ·51 পাওয়া গেল। এখন r= ·51 হইল X এবং Y variable—দ্বয়ের অনুবন্ধ সহগ।

26নং সূত্র দ্বারা r বাহির করিবার কালে নম্বরগুলির দূরত্ব (deviation) গুলিকে প্রকৃত Mean হইতে লওয়া হয়। তাই সেখানে deviation গুলি সর্বদা পূর্ণ সংখ্যায় না পাওয়ায় দশমিকের শরণাপন্ন হইতে হয়। তাই ইহাকে দীর্ঘ পদ্ধতি বলে। এই অসুবিধা দূর করিবার জন্য নম্বরগুলির দূরত্ব Mean হইতে না লইয়া Assumed Mean হইতে লওয়া হয়। কিন্তু তাহার জন্য যে ভুল হইবে তাহার সংশোধন প্রয়োজন। 27নং সূত্র দ্বারা সংক্ষিপ্ত পদ্ধতিতে সংশুদ্ধ r বাহির করা যায়। নীচে সূত্র এবং 20নং টেবলে ঐ পদ্ধতিতে r বাহির করার নিয়ম প্রদত্ত হইল।

$$r = \frac{\frac{\Sigma x^1 y^1}{N} - c_x\, c_y}{\sigma^1_x\, \sigma^1_y} \qquad \ldots\ldots\ldots\ldots (27)$$

যেখানে x^1 এবং y^1, Assumed Mean হইতে নম্বর গুলির দূরত্ব,

c_x, c_y, Assumed Mean লওয়ার জন্য সংশোধন,

এবং σ^1_x এবং σ^1_y দুইটি variable-এর অসংশোধিত SD.

ছাত্র	বিষয় (X)	বিষয় (Y)	x^1	y^1	x^{1^2}	y^{1^2}	x^1y^1
তাপস	85	65	5	5	25	25	25
সঞ্জয়	72	58	−8	−2	64	4	16
পথিক	75	55	−5	−5	25	25	25
কুমার	82	59	2	−1	4	1	−2
সৌমেন	75	63	−5	3	25	9	−15
সুদীপ্ত	80	56	0	−4	0	16	0
সুব্রত	83	62	3	2	9	4	6
জয়ন্ত	76	57	−4	−3	16	9	12
					168	93	67
					(Σx^{1^2})	(Σy^{1^2})	(Σx^1y^1)

$$AM_x = 80 \qquad AM_y = 60$$

$$c_x = -1{\cdot}5 \qquad c_y = -{\cdot}63$$

$$\sigma^1_x = \sqrt{\frac{168}{8} - 2{\cdot}25} \qquad \sigma^1_y = \sqrt{\frac{93}{8} - {\cdot}397}$$

$$= 4{\cdot}33 \qquad = 3{\cdot}35$$

$$r = \frac{\frac{67}{8} - (-1{\cdot}5)(-{\cdot}63)}{4{\cdot}33 \times 3{\cdot}35}$$

$$= \frac{7{\cdot}43}{14{\cdot}51} = {\cdot}51 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (27)$$

20নং টেবল : 8 জন ছাত্রলব্ধ শ্রেণীবিহীন নম্বরের সাহায্যে দুই বিষয়ের অনুবন্ধ সহগ নির্ণয় (সংক্ষিপ্ত পদ্ধতি)।

(b) Scatter-gram হইতেও আমরা Product Moment r বাহির করিতে পারি। নীচে 21নং টেবলে পদ্ধতি দেওয়া হইয়াছে। ধাপগুলি নীচে ক্রমান্বয়ে দেওয়া হইল।

(ক) পূর্ব বর্ণিত উপায়ে Scatter-gram-টি পূর্ণ করা হইল।

$$M_x = 52{\cdot}86 \qquad M_y = 44{\cdot}50$$

$$c_x = -{\cdot}33 \qquad c_y = {\cdot}50$$

$$c_x^2 = {\cdot}11 \qquad c_y^2 = {\cdot}25$$

$$\sigma^1_x = (1{\cdot}04) \times 10 \qquad \sigma^1_y = (1{\cdot}3) \times 5$$

$$r = \frac{\frac{\Sigma x^1y^1}{N} - c_xc_y}{\sigma^1_x\sigma^1_y} = \frac{\frac{28}{58} - (-{\cdot}33) \times {\cdot}5}{1{\cdot}04 \times 1{\cdot}3} = {\cdot}48$$

[যেহেতু $\sigma^1_x = 10\sqrt{\frac{69}{58} - {\cdot}11} = 10 \times (1{\cdot}04)$

এবং $\sigma^1_y = 5\sqrt{\frac{105}{58} - {\cdot}25} = 5 \times (1{\cdot}3)$]

P – test (X – Variable)

Q-test (Y-Variable)

	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	fy	y'	fy'	fy'2	Σx'y' +	Σx'y' −	Σx'	Σx'y'
55-59				(0) 1 0	(3) 1 3	(6) 1 6	3	3	9	27	9		3	9
50-54			(– 2) 1 –2	(0) 4 0	(2) 4 8	(4) 1 4	10	2	20	40	12	2	5	10
45-49		(– 2) 2 –4	(–1) 6 –6	(0) 6 0	(1) 2 2		16	1	16	16	2	10	–8	–8
40-44		(0) 2 0	(0) 6 0	(0) 6 0	(0) 2 0		16	0	0	0			–8	0
35-39		(2) 1 2	(1) 4 4	(0) 4 0	(–1) 1 –1		10	–1	–10	10	6	1	–5	5
30-34	(6) 1 6	(4) 1 4	(2) 1 2				3	–2	–6	12	12		–6	12
fx	1	6	18	21	10	2	58		29	105	41	– 13 (28)	–19	28
x'	–3	–2	–1	0	1	2								
fx'	–3	–12	–18	0	10	4 = –19								
fx'2	9	24	18	0	10	8 = 69								
Σy'	–2	–1	2	13	12	5 = 29								
Σx'y'	6	2	–2	0	12	10 = 28								

21নং টেবল : Scatter-gram এর সাহায্যে দুই Variable-এর r নির্ণয় (Product Moment Method).

(খ) ইহার পরে y^1 স্তম্ভটি পূরণ করিতে হইবে। f_y স্তম্ভের এক স্থানে—সাধারণতঃ মাঝামাঝি বরাবর 0 লইয়া উপরে 1, 2, 3 প্রভৃতি এবং নীচে $-1, -2, -3$ লেখা হইল। f_y এবং f_x এর যে ধাপ দুইটিতে 0 ধরা হইয়াছে সেই ধাপ বরাবর লাইন টানিয়া Scattergram-এর ভরকেন্দ্র দেখান হইয়াছে। পূর্ববর্ণিত নিয়ম অনুযায়ী fy^1, fy^{1^2} স্তম্ভদ্বয় এবং fx^1, fx^{1^2} সারিদ্বয় পূরণ করা হইল। উহাদের যোগফল বাহির করিয়া নীচে লেখা হইল।

(গ) এইবার Σx^1y^1 স্তম্ভটি পূরণ করিতে হইবে। ঐ স্তম্ভে '+' এবং '–' এর জন্য দুইটি বিভাগ করা হইয়াছে। প্রথমতঃ সর্বোপরি $y^1 = +3$ এবং সেই লাগোয়া দক্ষিণদিকের সর্বোচ্চ কোষ বরাবর $x^1 = +2$. সুতরাং গুণ করিলে $(+3)(+2) = +6$. উহাকে সর্বোচ্চ দক্ষিণ কোষের মাথায় লেখা হইয়াছে। উহার সহিত ঐ কোষে অবস্থিত '1' f-এর সহিত গুণ করিয়া নীচে কোষের কোণে লিখিত হইয়াছে। এইরূপ পরবর্তী বামকোষটিকে পূরণ করিতে হইবে। সেখানে $y^1 = +3$ এবং সেই কোষ বরাবর $x^1 = +1$. সুতরাং $x^1y^1 = (+3)(+1) = +3$. তাহাকে 1 এর উপর লিখিয়া পরে +3 কে ঐ f '1' এর সংগে গুণ করিয়া +3 দ্বিতীয় কোষের কোণে লেখা হইয়াছে। একই প্রক্রিয়ায় তৃতীয় কোষটি পূরণ করা হইয়াছে। এখন উপরের ঐ সারির তিনটি কোষের $+\Sigma x^1y^1$ হইবে $(+6+3+0) = +9$ এবং $-\Sigma x^1y^1$ স্তম্ভে কিছুই হইবে না। সুতরাং Σx^1y^1 স্তম্ভের '+' বিভাগে 9 লেখা হইল এবং '–' বিভাগে কিছুই লেখা হইল না। এইরূপে দ্বিতীয় সারির কোষগুলি পূর্ণ করা হইল। দেখা গেল $+\Sigma x^1y^1 = +12$ এবং $-\Sigma x^1y^1 = -2$. সেইজন্য '+' বিভাগে 12 এবং '–' বিভাগে 2 লেখা হইল। এই পদ্ধতি অবলম্বন করিয়া Scattergram-এর সমস্ত কোষ এবং Σx^1y^1 স্তম্ভটি পূরণ করা হইল। নীচে ঐ দুই বিভাগের যোগ করিয়া উহাদের বীজগাণিতিক যোগফল বাহির করা হইল। এখানে এই যোগফল হইয়াছে $[41+(-13)] = 28$.

(ঘ) ইহার পরের স্তম্ভটি হইল Σx^1. এখানে প্রতি সারির f কে সেই বরাবর x^1 দ্বারা গুণ করিয়া তাহাদের যোগফলগুলিসারি বরাবর লিখিতে হয়। যেমন উপর হইতে তৃতীয় সারি। উহাতে ডানদিক হইতে f যথাক্রমে 2, 6, 6, 2 রহিয়াছে। ঐ f বরাবর x^1 গুণ করিলে $(2\times1)+(6\times0)+6(-1)+2.(-2) = -8$ পাওয়া যাইবে। তাই ঐ তৃতীয় সারি বরাবর Σx^1 স্তম্ভটিতে—8 লেখা হইয়াছে। এই ভাবে প্রতি সারির জন্য Σx^1 বাহির করিয়া নীচে যোগফল

লেখা হইল। এইরূপে নীচে Σy^1 এর সারিটি পূরণ করা হইয়াছে। এখানে Σx^1 স্তম্ভের যোগফল হইয়াছে -19 এবং Σy^1 সারির যোগফল হইয়াছে 29.

(ঙ) এর পর Σx^1y^1 স্তম্ভ ও সারি পূরণ করিতে হইবে। Σx^1 এবং Σy^1 গুণ করিলেই Σx^1y^1 স্তম্ভের প্রতিটি সংখ্যা পাওয়া যাইবে। এইরূপে সারির Σx^1 এবং Σy^1 গুণ করিয়া নীচের Σx^1y^1 সারি পূরণ করা যাইবে। স্তম্ভ এবং সারির Σx^1y^1 বাহির করিয়া নীচে লেখা হইল। এখানে দুই ক্ষেত্রেই 28 হইল।

এখানে লক্ষ্য রাখা দরকার যে Scatter-gram-এ r বাহির করিতে ভুল হইতেছে কিনা তাহা সংকেত দ্বারা স্থিরীকৃত হয়। যেমন

(i) Σfy^1 এবং Σy^1 স্তম্ভ দুইটির ফল একই হইবে।

(ii) Σfx^1 এবং Σx^1 সারি দুইটির ফল একই হইবে।

এবং (iii) প্রতি Σx^1y^1 স্তম্ভ এবং Σx^1y^1 সারির ফল একই হইবে।

এখন উপরোক্ত সকল কার্য সমাধা হইলে 27নং সূত্রানুযায়ী r বাহির করা যাইবে। এখানে $r = \cdot 48$.

D. Regression এবং তাহার দ্বারা ভবিষ্যৎবাণী

একটি প্রদত্ত variable-এর মাধ্যমে আমরা অন্য একটি variable-এর মান নির্দেশ করিতে পারি। 28 নং চিত্রে উহা দেখান হইয়াছে। যে ছাত্র P-test-এ 64 নম্বর পাইয়াছে সে Q-test-এ কি নম্বর পাইতে পারে? সঠিক না হইলেও আন্দাজ মত নম্বর বলা সম্ভবপর। P-test-এর 64 নম্বর 60-69 এই ধাপে অবস্থিত। এখন এই ধাপে Qতে যত ছাত্র পড়ে তাহাদের Meanই হইল ঐ ছাত্রের Q-test-এর সম্ভাব্য নম্বর। ঐ ধাপে Q-test-এ 10 জন ছাত্র আছে। ঐ স্তম্ভে Q-test-এর Mean হইল 45·5. সুতরাং যে ছাত্র P-test-এ 64 নম্বর পাইয়াছে, Q-test-এ তাহার 45·5 নম্বর পাওয়ার সম্ভাবনা রহিয়াছে। এইরূপে P-test-এ যদি কেহ 45 নম্বর পায়, Q-test-এ তাহারও সম্ভাব্য নম্বর বলা যাইতে পারে। P-test-এ 45 নম্বর 40-49 ধাপে আছে। ঐ ধাপে 18 জন ছাত্র আছে। এখন ঐ 18 জন ছাত্রের Q-test-এর Mean 40·06. সুতরাং P-test-এ যখন 45 নম্বর তখন Q-test-এর সম্ভাব্য নম্বর হইবে 40·06·

এইরূপে বিপরীতক্রমে প্রদত্ত Q-test-এ নম্বরের মাধ্যমে P-test-এর সম্ভাব্য নম্বর বলা সম্ভবপর। যেমন ধরা গেল একটি ছাত্র Q-test-এ 43 নম্বর

পাইয়াছে। Q-test-এ 43 নম্বর 40-44 ধাপে আছে। এখন ঐ সারি বরাবর যে 16 জন ছাত্র আছে তাহাদের P-test-এ Mean নম্বর হইবে তাহার P-test-এর সম্ভাব্য নম্বর। ঐ 16 জন ছাত্রের Mean হইল 49·5. সুতরাং যে ছাত্র Q-test-এ 43 নম্বর পাইয়াছে তাহার P-test-এ সম্ভাব্য নম্বর হইবে 49·5. এইরূপে যে কোন Q-test-এর নম্বরের মাধ্যমে P-test-এ সম্ভাব্য নম্বরের ভবিষ্যৎবাণী করা সম্ভব।

Regression line-দ্বয়ের মাধ্যমে সোজাসুজি ভাবে প্রদত্ত একটি variable এর জন্য অন্য একটি variable নির্ণয় করিতে পারা যায়। তাহার জন্য সমীকরণের প্রয়োজন হয়। ঐ সমীকরণে একটি মান বসাইয়া নির্ণেয় অন্য একটির মান বাহির করা যায়।

Regression line-দ্বয়ের সমীকরণ দুইটি এমন দুইটি সরলরেখা নির্দেশ করিবে যাহারা স্তম্ভ এবং সারির Mean বিন্দুসমূহ নির্দেশক দুইটি 'best fit' সরলরেখা। নীচে প্রথম regression line-এর সমীকরণ দেওয়া হইল।

$$\bar{y} = r\,\frac{\sigma_y}{\sigma_x}\,.\,x \cdots\cdots\cdots\cdots\cdots\cdots \qquad (28)$$

যখন $\bar{y}$ এর অর্থ হইল সম্ভাব্য y-এর মান। ইহা M_y হইতে y দূরত্বে প্রকাশিত গড় মান বলিয়া y-এর মাথায় '–' চিহ্ন দেওয়া হয়।

r=X এবং Y variable-এর অনুবন্ধ সহগ

σ_y এবং σ_x = দুই variable-এর SD.

x = প্রদত্ত x-এর মান। ইহা M_x হইতে x দূরত্বে প্রকাশিত।

এখানে মনে রাখা দরকার যে 28 নং সূত্রের x এবং y প্রকৃত নম্বরের সমীকরণ নয়—Mean হইতে দুইটি variable-এর দূরত্বের সমীকরণ।

এখন ধরা গেল দুইটি variable-এর $r = \cdot75$, $\sigma_y = 5\cdot00$, $\sigma_x = 15\cdot00$, $M_y = 24\cdot00$ এবং $M_x = 150$. r, σ_y এবং σ_x-এর উক্ত মান 28 নং সূত্রে স্থাপন করিলে আমরা পাই যে

$$\bar{y} = \cdot75\tfrac{5}{15}\,.\,x = \tfrac{1}{4}\,x$$

স্পষ্টই বুঝা যাইতেছে যে y, x-এর তুলনায় এক চতুর্থাংশ পরিবর্তনের সম্মুখীন হয়। যদি কোন ছাত্র Mean হইতে 18 নম্বর বেশী পায়—অর্থাৎ 150+18=168, তবে তার y এর পরিবর্তন হইবে 18 এর $\frac{1}{4} = 4\cdot5$ অর্থাৎ তার Y-এ সম্ভাব্য মান হইবে (24+4·5)=29·5.

এইরূপে আমরা প্রদত্ত y-এর জন্য x-এর মানের ভবিষ্যৎবাণী করিতে পারি। তখন সমীকরণটি হইবে নিম্নরূপ (29 নং সূত্র)।

$$\bar{x} = r\,\frac{\sigma_x}{\sigma_y}.\,y \quad \ldots\ldots\ldots\ \ldots\ldots \qquad (29)$$

উপরের r, σ_x এবং σ_y-এর মানগুলি 29 নং সূত্রে স্থাপন করিলে

$$\bar{x} = \cdot75\tfrac{15}{5}y = 2\cdot25y.$$

অর্থাৎ y তাহার Mean হইতে যত দূরে যাইবে x তাহার Mean হইতে 2·25 গুণ দূরে যাইবে। অর্থাৎ $y=13$ হইলে $(Y=24+13=37)$, x-এর মান হইবে $13\times2\cdot5=32\cdot5$, অর্থাৎ X হইবে $150+32\cdot5=182\cdot5$.

28নং সূত্রের ন্যায় 29নং সূত্রটিও দুইটি Variable-এর Mean হইতে দূরত্বের সমীকরণ। Regression সমীকরণে (28নং এবং 29নং সূত্র) লিখিত $r\,\frac{\sigma_y}{\sigma_x}$ এবং $r\,\frac{\sigma_x}{\sigma_y}$—দুইটিকে regression সহগ বলা হইয়া থাকে।

প্রশ্ন উঠিতে পারে যে ভবিষ্যৎবাণীর জন্য দুইটি সমীকরণের প্রয়োজন কেন হয়? কারণ এই যে প্রথমটিতে (28নং সূত্র) প্রদত্ত x-এর মাধ্যমে সম্ভাব্য y-এর মান পাইয়া থাকি। ইহা ঠিক y নহে। সম্ভাব্য y-গুলির গড়। এইরূপ দ্বিতীয়টির (29নং সূত্র) দ্বারা একটি প্রদত্ত মানের সাহায্যে অন্যটির সম্ভাব্য গড় মান বাহির করা হয়। যখন $r=1\cdot00$ হয়, তখন দুইটি সমীকরণ একই সমীকরণে পর্য্যবসিত হয়।

$$\bar{y} = r\,\frac{\sigma_y}{\sigma_x}.\,x = \frac{\sigma_y}{\sigma_x}\,x \text{ (যখন } r=1\cdot00) \quad \therefore\ \bar{y}\sigma_x = x\sigma_y$$

$$\text{আবার } \bar{x} = r\frac{\sigma_x}{\sigma_y}\,y = \frac{\sigma_x}{\sigma_y}\,y \text{ (যখন } r=1\cdot00) \quad \therefore\ \bar{x}\sigma_y = y\sigma_x$$

এক কথায় $r=\pm1\cdot00$ হইলে সরল রেখা দুইটি উপর্য্যুপরি অবস্থিত থাকে। যখন নম্বর দ্বারা প্রকাশ করা হয় তখন regression সমীকরণ দুইটি নিম্নরূপ ধারণ করে।

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}\,(X - M_x)$$

$$\text{অথবা, } \bar{Y} = M_y + r\frac{\sigma_y}{\sigma_x}(X - M_x) \ldots\ldots\ldots\ldots\ldots\ldots \ldots \qquad (30)$$

$$\text{এবং } X - M_x = r\frac{\sigma_x}{\sigma_y}\,(Y - M_y)$$

$$\text{অথবা, } \bar{X} = M_x + r\frac{\sigma_x}{\sigma_y}(Y - M_y) \ldots\ldots\ldots\ldots\ \ldots\ldots \qquad (31)$$

এখন পূর্বে উল্লিখিত σ_x, σ_y, r, M_x এবং M_y-এর মান 30নং সূত্রে স্থাপন করিলে আমরা পাই যে,

$\overline{Y}=24+\cdot75\frac{6}{18}(X-150)$

অর্থাৎ $\overline{Y}=\cdot25X-13\cdot5$

এখন X-এর মান 120 হইলে Y-এর সম্ভাব্য মান হইবে গড়ে 16·5. আবার এইরূপে ঐ মানগুলি 31নং সূত্রে স্থাপন করিলে

$\overline{X}=150+\cdot75\frac{18}{6}(Y-24)$

অর্থাৎ $\overline{X}=2\cdot25Y+96$

এখন Y-এর মান 12 হইলে X-এর সম্ভাব্য মান হইবে গড়ে 123.

উপরের উদাহরণ হইতে স্পষ্টই বুঝা যাইতেছে যে regression সমীকরণ হইতে ভবিষ্যৎ বাণী করা সম্ভবপর। কি মনোবিজ্ঞানে, কি শিক্ষায়, কি বৃত্তি নির্বাচনে regression সমীকরণ এই ভবিষ্যৎবাণী করিতে সক্ষম। নূতন ক্লাসে ছাত্র ভর্তি করিতে হইবে। পূর্ব বৎসরে সেশনের পূর্বে এবং পরে দেওয়া অভীক্ষার নম্বরের অনুবন্ধ সহগ যদি ·8 হয়, তবে তাহা হইতে যে Regression সমীকরণ তৈরী হইবে তাহা দ্বারা নূতন বৎসরের প্রথমে প্রযুক্ত সেই একই অভীক্ষার ফলের মাধ্যমে নূতন শ্রেণীতে ছাত্র নির্বাচন বা ছাত্রদের ভবিষ্যৎ সাফল্য সম্বন্ধে কিছুটা নির্দেশ দেওয়া যায়। শিক্ষা বা বৃত্তিতে এই ভবিষ্যৎবাণী একান্ত অপরিহার্য। ছাত্রগণ নিজ নিজ ঝোঁক, ক্ষমতা এবং আগ্রহ লইয়া শিক্ষা বা বৃত্তিতে নিযুক্ত হইলে ব্যক্তি তথা সমাজের উপকার সাধিত হইবে।

E. Regression সমীকরণ হইতে প্রাপ্ত ভবিষ্যৎবাণীর যাথার্থ্য ঃ

Regression সমীকরণ হইতে পূর্বে X বা Y-এর যে ভবিষ্যৎবাণী করা হইয়াছে তাহা সম্ভাব্য মান—সঠিক মান নহে। ঐ ভবিষ্যৎবাণী একটি গড় মান নির্দেশ করে। প্রকৃত মান উহা হইতে কিছু কম বা বেশী হইতে পারে। 28নং চিত্রে বলা হইয়াছে যে, যে ছাত্র P-test-তে 64 নম্বর পাইয়াছে, Q-test-এ তাহার 45·5 নম্বর পাওয়ার সম্ভাবনা রহিয়াছে। কিন্তু আমরা জানি যে Q-test-এ তাহার 35 হইতে 59 নম্বর পাওয়ার সম্ভাবনা আছে। Q-test-এ 35 হইতে 59 নম্বরের সীমার মধ্যে যে 10 জন ছাত্র আছে তাহাদের গড় 45·5 বাহির করিয়া তাহার Q-test-এ নম্বরের ভবিষ্যৎবাণী করিয়াছিলাম। সুতরাং ঐ ভবিষ্যৎবাণীর মধ্যে কিছু ভ্রম রহিয়া যাইতেছে। আমরা তাহার

Standard Error বাহির করিব এবং ঐ মানের উপর কতটা ভরসা করিব তাহা এখন নির্ণয় করা হইবে।

$\bar{Y}$ নম্বরের S E কে $\sigma_{(est\ Y)}$ এবং $\bar{x}$ নম্বরের S E কে $\sigma_{(est\ X)}$ বলা হইয়া থাকে। যখন প্রদত্ত X মানের মাধ্যমে $\bar{Y}$ বাহির করা হয় তখন $\bar{y}$ এর SE হইবে নিম্নরূপ।

$$\sigma_{(est\ Y)} = \sigma_y\sqrt{1-r^2} \quad \ldots\ldots\ \ldots \qquad (32)$$

যখন σ_y = Y দলের SD

এবং r = X এবং Y variable-দ্বয়ের অনুবন্ধ-সহগ।

21 নং টেবলে কোন ছাত্র P-test-এ যদি 54 নম্বর পায় তবে তাহার Q-test-এর ফল বাহির করিতে হইলে regression সমীকরণটির দ্বারস্থ হইতে হইবে।

$$\bar{Y} = \cdot 48\ \frac{6\cdot 5}{10\cdot 4}(X - 52\cdot 86) + 44\cdot 50$$

$$\therefore\quad \bar{Y} = \cdot 3X + 28\cdot 64$$

এখন X-এর মান ঐ সমীকরণে স্থাপন করিলে Q-test-এ তাহার নম্বর দাঁড়ায় 44·84 বা প্রায় 45. তাহা হইলে যে ছাত্র P-test-এ 54 নম্বর পাইয়াছে তাহার Q-test-এ সম্ভাব্য নম্বর হইবে 45. এখন আমরা এই সম্ভাব্য নম্বরের যাথার্থতা বাহির করিব। 32 নং সূত্র দ্বারা আমরা জানি যে,

$$\sigma_{(est\ Y)} = 6\cdot 5\sqrt{1 - \cdot 48^2}$$

$$= 5\cdot 72$$

এখন আমরা বলিতে পারি যে, যে ছাত্র P-test-এ 54 নম্বর পাইয়াছে তাহার Q-test-এ সম্ভাব্য নম্বর হইবে 45, এবং এই সম্ভাব্য নম্বরের SE হইতেছে 5·72. অর্থাৎ 100 জনের মধ্যে প্রায় 68 জন উক্ত সম্ভাব্য নম্বর পাইবে এবং তাহাতে $\pm 1\sigma_{(est\ Y)}$ বা ± 5·72 এর বেশী ভুল হইবে না।

এইরূপে আমরা সম্ভাবনা আরও বাড়াইয়া দিতে পারি—যদিও ভুলের সীমা ও সংগে সংগে বৃদ্ধি পাইবে। যেমন 100 জনের মধ্যে 95 জন ছাত্র ঐ সম্ভাব্য

নম্বর পাইবে—এবং তাহাতে ভুল হইবে $\pm 2\,\sigma_{(est\ Y)}$ বা $\pm$ 11·44. আবার 100 জনের মধ্যে প্রায় 100 জনই (99·7) ঐ সম্ভাব্য নম্বর পাইবে এবং তাহাতে ভুল হইবে $\pm 3\sigma_{(est\ Y)}$ বা $\pm$ 17·16.

এইরূপে আমরা X-এর সম্ভাব্য নম্বরের SE বাহির করি। তখন সূত্র দাঁড়ায় নিম্নরূপ।

$$\sigma(\text{est x}) = \sigma_x\sqrt{1-r^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(33)$$

$S.E_{(est)}$ প্রয়োগ করিতে গেলে তিনটি সর্তের দিকে লক্ষ্য রাখিতে হইবে।

(ক) প্রথমতঃ regression line যেন সরলরৈখিক চিত্র হয়। সাধারণতঃ শিক্ষা বা মানসিক গুণাবলী সরলরৈখিক regression line প্রদান করে। কিন্তু এমন অনেক সামাজিক ব্যক্তিত্ব সংক্রান্ত গুণ আছে যাহারা তদ্রূপ নয়। তখন $SE_{(est)}$ বাহির করার কোন মূল্য থাকে না।

(খ) দ্বিতীয়তঃ x এবং y variable-এর বিস্তৃতি Normal বা Normal-এর কাছাকাছি থাকা দরকার। Distribution-এর কোনটির অষথা Skewness বা Kurtosis থাকিলে SE(est) প্রয়োগ ব্যর্থ হইবে।

(গ) তৃতীয়তঃ Scattergram-এর প্রত্যেক স্তম্ভ এবং সারির বিস্তৃতি সমান হওয়া উচিত।

উল্লিখিত তিনটি সর্ত পূর্ণ হইলে SE(est) ব্যক্তি বিশেষের ভবিষ্যৎ বাণীর যথার্থ সীমা প্রদান করে। SE(est), দুই বিস্তৃতির SD এবং দুইটি variable-এর r-এর উপর নির্ভরশীল। r যদি 1·00 হয় তবে SE(est)-এর মান শূন্য হইবে। ইহার অর্থ এই যে প্রদত্ত variable-এর (x) মাধ্যমে যে অজানা Variable-টি (y) বাহির হইবে তাহা ভুলশূন্য। যদি r=·00 হয় তবে SE(est)-এর মান SD-এর সমান হইবে। অর্থাৎ প্রতি ব্যক্তির $\overline{Y}$-এর মান হইবে M_y. ব্যক্তির সম্ভাব্য মান distribution-এর যে কোন স্থানে থাকিতে পারে। যদি σ_y ছোট এবং r বড় হয়, তবে শঙ্কাবিহীন ভাবে $\overline{Y}$-এর মান বাহির করা যায়। ইহার বিপরীত হইলে $\overline{Y}$ মানের উপর নির্ভর করা যায় না।

$\overline{Y}$ বা $\overline{X}$-এর মান বাহির করিয়া SE(est) বাহির করা উচিত। ইহাদের উপরেই $\overline{Y}$ বা $\overline{X}$-এর মানের যাথার্থ নির্ভর করে।

Exercise VII

1. Find the correlation between the two sets of Scores given below, using the ratio method.

a)

X	Y
22	11
8	5
19	6
32	8
13	2
24	5
22	4
35	1
18	7
13	10

b)

X	Y
2	10
20	4
25	11
14	6
11	2
2	9
38	17
16	6
14	4
23	25

2. Calculate the coefficient of correlation between the two sets of scores given below, using the ratio method.

a)

X	Y
11	24
5	22
6	44
8	72
2	25
5	30
4	38
1	54
7	37
10	61

b)

X	Y
10	29
4	50
11	76
6	40
2	32
9	61
17	56
6	61
4	17
25	61

3. Prepare a scatter diagram for the correlation of the scores given below. Compute a Pearson r and state conclusions.

X	Y	X	Y	X	Y	X	Y
24	29	56	45	67	48	29	49
22	40	42	48	58	48	50	55
44	36	30	25	57	33	76	43
72	32	42	48	49	47	40	38
25	46	28	28	87	52	32	56
30	47	32	40	14	48	61	45
38	49	24	37	38	46	56	67
54	53	42	58	32	33	61	42
37	51	54	54	52	40	17	44
61	50	42	44	60	49	61	48

4. Compute the coefficient of correlation in the table below.

X →

Y ↑	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	fy
135-139								1	1
130-134				1	1		1		3
125-129				1		2	1		4
120-124			1	4	4	6	2		17
115-119			7	5	7	2	1		22
110-114	1	4	2	9	4	2			22
105-109	1	1	2	5	1				10
100-104	1	3		1	1				6
95-99		2							2
fx	3	10	12	26	18	12	5	1	87

5. In the following correlation table compute the coefficient of correlation.

School Marks	84 and lower	85-	90-	95-	100-	105-	110-	115-	120-	125-over	Totals
90 and over				3	3	15	12	9	9	5	56
85-89				8	17	15	24	13	6	6	89
80-84			4	6	22	21	20	10	5	1	89
75-79			7	25	33	23	10	7	4		109
70-74		4	10	18	14	22	12	1	1		82
65-69	1	3	3	12	7	8	8	1			43
60-64			2	5	3	1	1				12
Totals	1	7	26	77	99	105	87	41	25	12	480

6. Write out the regression equations in score form for the correlation table in example 4, Exercise VII.

(a) Compute $\sigma_{\text{(est Y)}}$ and $\sigma_{\text{(est X)}}$

(b) What is the most probable (score in Y) of a boy whose score in X is 76 ? 66 ? 92 ? What is the most probable score in X of a boy whose score in Y is 132 ? 116 ? 100 ?

7. In example 4, page 148, find the most probable grade whose I Q is 100 (data from example 5, Exercise—VII) ? What is the $\sigma_{\text{(est)}}$ of this grade ?

Answers.

1. (a) r = − ·16
   (b) r = + ·47
2. (a) r = ·18
   (b) r = ·49
3. r = ·25
4. r = ·61
5. r = ·46

অষ্টম অধ্যায়

Mean এবং অন্যান্য পরিসংখ্যানের নির্ভরযোগ্যতা নির্ণয়।

সাধারণতঃ পরিসংখ্যান যে ক্ষেত্র হইতে প্রাপ্ত হয় সেই ক্ষেত্র বড় সংকীর্ণ। ছোট ছোট দলের উপর অভীক্ষা প্রয়োগ করিয়া আমরা সূত্র নির্ণয় করি। সেখানেই আমাদের ভ্রান্তির সম্মুখীন হইতে হয়। ছোট দলের ফলের উপর নির্ভর করিয়া তাহাকে ব্যাপক ক্ষেত্রে প্রয়োগ করিতে হইলে ভুলের সম্ভাবনা তো থাকিবেই। ৪ বৎসর বয়স্ক একটি ছোট ছাত্রদলের উপর প্রযুক্ত কোন অভীক্ষার ফলকে যদি সমগ্র ৪ বৎসর বয়সের ছাত্রের ফলের প্রতিভূ হিসাবে গণ্য করা হয়, তাহা হইলে আমাদের কিছুটা ভুল হওয়ার সম্ভাবনা থাকে। কিন্তু ৪ বৎসর বয়স্ক পৃথিবীর সব ছাত্র সংগ্রহ করা বা তাহাদের উপর অভীক্ষা প্রয়োগ করা সম্ভবপর নয় এবং অপ্রয়োজনীয়ও বটে। অপেক্ষাকৃত ছোট দলের (Sample) মাধ্যমে বৃহত্তম ব্যাপক দলের (population) আন্দাজ কিরূপে করা যায় বা সেই আন্দাজে কতটুকু ভুল আছে, বা কোন্ সর্ত্তে সেই আন্দাজ নির্ভরযোগ্য বা আন্দাজ মোটেই সম্ভবপর কিনা—এই অধ্যায়ে তাহার আলোচনা হইবে। তবে এ কথা ঠিক যে পরিসংখ্যানে সম্ভাবনা—নীতির একটি বিশিষ্ট স্থান আছে এবং তাহার প্রয়োগ কয়েকটি বিশেষ সর্ত্তাধীনে পরিচালিত হয়। আর যদি সেই সকল সর্ত্ত প্রতিপালিত না হয়, যত সতর্কতার সংগে নির্ভুলভাবে ভবিষ্যৎ বাণী করার চেষ্টা করি না কেন—তাহার ভিতর Standard Error এত উগ্রভাবে দেখা দিবে যে গোটা ভবিষ্যৎবাণীর অস্তিত্বই সন্দেহ পূর্ণ হইবে এবং উহা আমাদিগকে ভুল পথে পরিচালিত করিবে। তবে উক্ত সর্তগুলি মানিয়া চলিলে আমরা কোন ধারণা সৃষ্টি করিতে পারি—কোন সিদ্ধান্তে আসিতে পারি—যদিও সেই সিদ্ধান্ত সম্পূর্ণ নির্ভুল নয়—তাহার মধ্যে সম্ভাবনার ছাপ বিদ্যমান। নিঃসংকোচে এবং নির্ভয়ে সেই কথা স্বীকার করিয়া আমরা ভবিষ্যৎ বাণী করার পথে অগ্রসর হইব। তাহার পূর্বে Sampling সম্বন্ধে সাধারণ ভাবে কিছু আলোচনার দরকার।

Random Sampling : যখন বৃহত্তম দল হইতে একটি ক্ষুদ্র দল লইয়া তাহাদের উপর পরিসংখ্যান পদ্ধতি প্রয়োগ করা হয়—তখন ঐ ক্ষুদ্র দলটিকে

Random উপায়ে বাছাই করা দরকার। কারণ ঐ ক্ষুদ্র দলটিই হইবে সমগ্রের প্রতিভূ। এখন Random Sampling কাহাকে বলে? যখন Population-এর মধ্য হইতে একটি দল এমন ভাবে বাছাই হয় যে Population-এর প্রতিটি ব্যক্তির বাছাই হওয়ার সমান সুযোগ থাকিবে তখন সেই বাছাইকে বলা হইবে Random sampling এক জনের বাছাই-এর সংগে অপরের বাছাই হওয়ার কোন সম্পর্ক থাকিবে না। সুপরিকল্পিত ও সুনিয়ন্ত্রিত লটারী পদ্ধতি, মুদ্রাক্ষেপণ পাশাচালনা প্রভৃতি যে সকল পদ্ধতি সম্ভাবনার নিয়ম মানিয়া চলে তাহাদের সাহায্যে Random sampling করা যাইতে পারে।

Biased Sampling :—Biased Sample ভুলে পরিপূর্ণ। এই পদ্ধতিতে বাছাই-এর সময় কিছু ব্যক্তির উপর পক্ষপাতিত্ব আসে। বাছাই-এর সময় সবাইকে সমান সুযোগ দেওয়া হয় না। অষ্টম শ্রেণীর জন্য একটি অভীক্ষা প্রস্তুত করিবার সময় যদি প্রাথমিক পরীক্ষা বালীগঞ্জ সরকারী বিদ্যালয় বা হেয়ার স্কুলের ছাত্র অথবা বেলতলা বালিকা বিদ্যালয়ের ছাত্রীদের মধ্যে সীমাবদ্ধ রাখা হয়—তাহা হইলে Sample-টি Biased হইবে। তাই দল বাছাই-এর সময় যেন তাহা Biased Sampling-এর দোষে দুষ্ট না হয় তাহার দিকে বিশেষ নজর দিতে হইবে।

Stratified Sample : Biased Sampling-এর ত্রুটী অপনোদন এবং প্রতিভূ দল বাছাই-এর জন্য একটি সাধারণ নিয়ম অনুসরণ করা হয়। ইহা Stratified Sampling. এই পদ্ধতিতে সমগ্র Population-কে কয়েকটি Strata-তে ভাগ করা হয় এবং সেই বিভাগ হইতে ছোট ছোট দল লইয়া গঠিত দলকে সমগ্রের প্রতিভূ বলিয়া ধরা হয়। মনে করা যাউক যে সরকার কোন একটি বিশেষ বিধির প্রচলন চান এবং পরে তাহাকে আইনে পরিণত করিতে চান। তাহা হইলে আইন প্রণয়নের পূর্বে সমগ্র লোকের সেই প্রসংগে মতামত প্রয়োজন। সমাজে বিভিন্ন প্রকার বিভাগ—যেমন বৃত্তিভোগী, ব্যবসায়ী, সরকারী চাকুরে, অর্ধদক্ষ কারিগর, অদক্ষ কারিগর প্রভৃতি বিদ্যমান। আবার ঐ সকলের মধ্যে পুরুষ ও মহিলার বিভাগও ধরা যাইতে পারে। এ ছাড়া সহুরে—গ্রামীন; স্থানীয়—ভিনদেশীয়, শিক্ষিত-অশিক্ষিত, প্রভৃতি বিভাগও আছে। উক্ত মত গ্রহণের নিমিত্ত কোন্ কোন্ বিভাগের প্রভাব তাহার উপর বিদ্যমান তাহা স্থির করিয়া Sample বাছিতে হয়। এখন কি সর্তে ঐ Sample-এ লোক নির্বাচিত হইবে? সমাজে ঐ বিভাগগুলির মোট পরিমাণ স্থির করা আছে; তাহাদের

আনুপাতিক সংখ্যা লইয়া সেই অনুযায়ী Sample-এ লোক বাছাই হয়। শেষোক্ত বাছাই Random Sampling দ্বারাই সম্পন্ন করা হয়। Population-এর Random Sample পূর্বে না করিয়া Strata-তে বিভক্ত করিয়া Random sampling করিলে যে ফল ভাল হয় ও Sample বাছাই যে অপেক্ষাকৃত খাঁটি প্রতিভূ প্রদান করে—তাহা অভিজ্ঞতা-প্রসূত ও প্রয়োগসিদ্ধ। বুদ্ধি বা শিক্ষাগত অভীক্ষা তৈরীর সময়ও এই পদ্ধতি অবলম্বন করা যাইতে পারে।

Purposing Sample : অনেক সময় দেখা যায় যে কোন একটি বিশেষ দলের উপর সমাজের লোকের মতামত নির্ভর করিতেছে। সেরূপক্ষেত্রে অন্যান্য দলকে বাদ দিয়াই ঐ বিশেষ দলগুলি হইতে Sample বাছাই করা হয়। ইহাই উদ্দেশ্যমূলক Sampling. কিন্তু এই পদ্ধতিতে ভুলের সম্ভাবনা বিদ্যমান। কারণ কোন নূতন মত গ্রহণের সময় ঐ বিশেষ দলের মতামত Population-এর প্রতিভূ নাও হইতে পারে। তাই এর প্রয়োগ সীমাবদ্ধ।

Incidental Sampling : অনেক সময় কাজের সুবিধার জন্য হাতের কাছে যে নমুনা পাওয়া যাইতেছে তাহাকেই প্রতিভূ নমুনা হিসাবে গ্রহণ করা হয়। ইহাকে Incidental Sample বলে। কিন্তু এই প্রকারে প্রাপ্ত কোন সিদ্ধান্ত বৃহত্তর ক্ষেত্রে প্রয়োগ করিবার কালে যথেষ্ট সতর্কতা অবলম্বন করিতে হয়।

B. গড়ের নির্ভরযোগ্যতা :

নমুনা দল : Population হইতে সমান সংখ্যা বিশিষ্ট ক্ষুদ্র ক্ষুদ্র Sample বাছাই করা হয়। মনে রাখিতে হইবে যে একটি Sample বাছাই-এর পর তাহাদিগকে আবার সমগ্র দলে মিশাইয়া দিয়া আর একটি Sample-এর বাছাই চলে। পূর্ব Sample-এর অনুপস্থিতিতে নূতন Sample-এর বাছাই চলে না। কারণ Random Sample-এর নিয়মই হইতেছে যে প্রতি ব্যক্তির বাছাই হওয়ার সমান সম্ভাবনা রহিয়াছে। Sample-এর সংখ্যা যত বাড়িবে, Sample সমূহের Mean, Population Mean-এর কাছে যাইবে। Sample-দের Mean-এর Mean-ই Population Mean. সুতরাং Sample-এ যদি একজন করিয়া ব্যক্তি থাকে তবে Sample Mean-সমূহের বিস্তৃতি এবং Population-এর বিস্তৃতি একই হইবে। কিন্তু Population Mean ধারণা-প্রসূত। বাস্তবক্ষেত্রে ইহা পাওয়া সম্ভবপর নয়; তাই আমাদিগকে কোন

Sample Mean-এর মাধ্যমে Population Mean-এর আন্দাজ করিতে হয়।

1. Mean-এর Standard Error (SE_M বা σ_M)ঃ

(ক) সমান সংখ্যাবিশিষ্ট (Population হইতে) বাছাই Sample-গুলির Mean-দিগকে লেখচিত্রের সাহায্যে প্রকাশ করা হয়। এই Sample Mean-গুলির বিস্তৃতি Population Mean-এর কিছুটা ধারণা দেয়। Sample Mean-কে যদি Population Mean হিসাবে ধরা হয় তাহা হইলে কিছু ভুল হয়। ঐ ভুলকে Sample Mean-গুলির SD-দ্বারা সূচিত করা হয়। ইহাকে Mean-এর Standard Error বা SE_M বা σ_M নামে আখ্যা দিয়া Standard Deviation হইতে পৃথক করা হয়।

কোন Mean-এর SE_M বা σ_M বাহির করিতে হইলে দুইটি রাশির দরকার হয়। যথা—(a) Population Mean-এর SD বা $\bar{\sigma}$ এবং (b) সেই Sample-এর জনসংখ্যা বা N.

$$\sigma_M = \frac{\bar{\sigma}}{\sqrt{N}} \qquad \cdots\cdots\cdots\cdots(34)$$

ইহা হইতে এই সিদ্ধান্তে আসা যায় যে σ যত বেশী হইবে σ_M তত বেশী হইবে এবং N বেশী হইলে ভুলের পরিমাপও কমিয়া যাইবে।

কিন্তু Population Mean-এর SD বা $\bar{\sigma}$ পাওয়া সম্ভবপর নহে। সেই জন্য নিম্নলিখিত সূত্রের সাহায্যে σ_M বাহির করা হয়।

$$\sigma_M = \frac{\sigma}{\sqrt{N-1}} \qquad \cdots\cdots\cdots\cdots(35)$$

(এখানে σ =Sample-এর SD)

(খ) Mean-এর Standard Error-এর ব্যাখ্যা

উদাহরণের সাহায্যে SE_M —এর ব্যাখ্যা করা হইতেছে। মনে করি 50 জন ছাত্র-ছাত্রীর উপর প্রযুক্ত শিক্ষাগত কোন বিষয়ে অভীক্ষার ফল Mean= 50 এবং $\sigma = 10{\cdot}5$ সুতরাং $\sigma_M = 1{\cdot}5$ এখন Population Mean হইতে এই রকম Sample Mean কত দূরে অবস্থিত তাহাই জিজ্ঞাস্য। আমরা Population Mean কত জানি না ; তবে $\sigma_M = 1{\cdot}5$ এর দিকে লক্ষ্য

রাখিয়া একথা নিঃসন্দেহে বলা যায় যে প্রায় দুই-তৃতীয়াংশ সময় 50 জন লইয়া গঠিত Sample Mean-গুলি, Population Mean-এর দুইদিকে 1·5 এককের বেশী দূরে থাকিবে না। কারণ আমরা ধরিয়া লইয়াছি যে 50 জনকে লইয়া গঠিত Sample Mean-গুলির বিস্তৃতি Normal curve-এর নিয়ম মানিয়া চলে। উপরোক্ত সিদ্ধান্তে অন্য আর একটি তথ্যের নিশানা দেয়। আমরা উহা হইতে অন্য সিদ্ধান্তও করিতে পারি। যেমন শতকরা প্রায় 68 ভাগ সম্ভাবনা রহিয়াছে যে Population Mean, Sample Mean হইতে সর্বাপেক্ষা বেশী এক σ_M বা 1·5 একক দূরে আছে। আবার শতকরা প্রায় 94 ভাগ সম্ভাবনা আছে যে Population Mean, Sample Mean হইতে সর্বাপেক্ষা বেশী 1·96 σ_M বা 3·0 একক দূরে থাকিতে পারে এবং প্রায় 99% সম্ভাবনা আছে যে সর্বাপেক্ষা বেশী 2·58 σ_M বা 3·87 একক দূরে থাকিতে পারে। তাই এই ভবিষ্যৎ বাণীর আস্থার মাত্রা (confidence level) তিনবারে তিন প্রকার হইয়াছে। দ্বিতীয়বারে 5% বা ·05 মাত্রা এবং তৃতীয় বারে 1% বা ·01 মাত্রায় ভবিষ্যৎ বাণী হইয়াছে। তাহা ছাড়া আমরা আস্থার যে কোন মাত্রা গ্রহণ করিতে পারি। তবে সাধারণতঃ উক্ত দুই মাত্রাতেই ভবিষ্যৎবাণী করা হইয়া থাকে।

2 . Median-এর নির্ভর যোগ্যতা

উক্ত প্রকারেই Sample-এর Mdn. বাহির করা হয়। কিন্তু Sample Mdn-গুলির বিস্তৃতি, Sample Mean-গুলির বিস্তৃতি অপেক্ষা শতকরা প্রায় 25 ভাগ বেশী। অবশ্য Population-এর বিস্তৃতি Normal হইতে হইবে। সেইজন্য Sample Mdn-এর Standard Error বা

$$\sigma_{Mdn.} = \frac{1{\cdot}253\sigma}{\sqrt{N}} \qquad \ldots\ldots\ldots(36)$$

$\sigma_{Mdn.}$ —এর ব্যাখ্যা σ_M —এর ন্যায় হইয়া থাকে। তাই তাহার পুনরালোচনা নিষ্প্রয়োজন।

3. অনুবন্ধ সহগের নির্ভরযোগ্যতা ঃ

অন্যান্য পরিসংখ্যানের ন্যায় অনুবন্ধ সহগ বা r-এর Standard Error আছে। মনে করা যাক Population $\bar{r}$ = 50. এখন যদি Population

হইতে সমান সংখ্যা বিশিষ্ট 50 জোড়া Sample বাছিয়া তাহাদের প্রতি জোড়ার r বাহির করা হয়, তবে দেখা যাইবে যে Population r-এর দুই পাশে তাহারা ভিড় জমাইয়াছে। এখন এই সকল Sample r-গুলির বিস্তৃতি কেমন হইবে? r-এর বিস্তৃতি সব সময় সমান নয়। ইহা r-এর মান এবং Sample-এর জনসংখ্যার উপর নির্ভরশীল। পূর্বেই বলা হইয়াছে যে r-এর মান $-1\cdot0$ হইতে $+1\cdot0$-এর মধ্যে অবস্থিত থাকে। যেহেতু Population r-কে ঐ সীমার মধ্যে থাকিতে হইবে, তাই Sample r ধনাত্মক হইলে r-এর বিস্তৃতি ধনাত্মক Skewed হইবে এবং বিপরীত ক্রমে, Sample r ঋণাত্মক হইলে r-এর বিস্তৃতি ঋণাত্মক Skewed হইবে। কেবলমাত্র r-এর মান যখন শূন্য হইবে, তখনই কেবল r-গুলি ঐ শূন্যের উভয় পার্শ্বে সদৃশ আকারে বিস্তৃত থাকিবে।

r-এর Standard Error বা σ_r নিম্নের সূত্র সাহায্যে পাওয়া যাবে।

$$\sigma_r = \frac{1-r^2}{\sqrt{N-1}} \qquad \cdots\cdots(37)$$

এখানে r-এর পরিবর্তে $\bar{r}$ লিখিলে আরও শুদ্ধ হইত। কিন্তু $\bar{r}$ আমাদের জানার বাহিরে। তাই r-কে লইয়াই আমাদিগকে সন্তুষ্ট থাকিতে হয়। আবার এটাও ঠিক যে r এবং N-এর মান বেশী হইলে ভুলের সম্ভাবনা কমিয়া যায়। যদি $r=\cdot80$ এবং $N=50$, এখন $\sigma_r=\cdot05$. অন্যান্য পরিসংখ্যানের মত ইহার সাহায্যে Population r বা $\bar{r}$ সম্বন্ধে একটা ধারণা দেওয়া হইতেছে। শতকরা প্রায় 95 ভাগ আস্থার সংগে বলা যায় যে $\bar{r}$, $r\pm1\cdot96\sigma_r$ বা $\cdot80\pm 1\cdot96\times\cdot05$ অর্থাৎ ·70 হইতে ·90-এর মধ্যে থাকিবে। সেই রকম শতকরা প্রায় 99 ভাগ আস্থার সংগে বলা যায় যে $\bar{r}$, $r\pm2\cdot58\sigma_r$ বা ·67 হইতে ·93-এর মধ্যে থাকিবে। প্রথমটিতে আস্থার মাত্রা ·05 এবং শেষোক্ত ক্ষেত্রে উহা ·01 হইবে।

r-এর নির্ভর যোগ্যতা (significance) বাহির করিতে হইলে σ_r বাহির করিতে হয়। সাধারণতঃ যদি σ_r-এর মানের তিন গুণ r-অপেক্ষা কম বা r-এর সমান হয় তাহা হইলে r নির্ভরযোগ্য—বেশী হইলে বুঝিতে হইবে যে r নির্ভর-যোগ্য নহে।

Exercise—VIII

1. Compute the standard errors of the means for the Data given below, and interpret your results. Determine

confidence limits at ·05 and ·01 levels. State the confidence intervals.

Measures	Boys	Girls
N	95	164
M	21·1	23·0
σ	5·2	4·8
Q	3·1	2·92
Mdn	20·6	23·4

2. Calculate the standard errors of the means for the Data given below, and interpret your results.

Measures	Adult	Child
M	62·5	59·61
σ	11·32	8·76

3. Compute the standard errors of the medians for the data given in problem 1, Exercise—VIII, and interpret your results.

4. Compute the standard errors of the standard deviations in the problems 1 & 2, Exercise—VIII, and interpret your results.

5. Compute the standard errors of the co-efficients of correlation in the problems 1, 2, 3, 4 and 5, Exercise—VII.

নবম অধ্যায়

পরীক্ষিত ধারণার সত্যতা যাচাই পদ্ধতি ঃ

পরীক্ষা এবং পরিসংখ্যান পদ্ধতি অঙ্গাঙ্গীভাবে জড়িত। পরীক্ষাকে সুষ্ঠু-ভাবে পর্যবেক্ষণ করিতে হয় এবং তাহা হইতে পরিসংখ্যানের উদ্ভব হয়। আমরা সংখ্যায় প্রকাশিত উক্ত পরিমাপের নির্ঘণ্ট তৈয়ারী করি—তাহার নির্ভরযোগ্যতা যাচাই করি। পরিশেষে উহা হইতে কোন বিশেষ সিদ্ধান্তে আসি এবং ভবিষ্যৎবাণী করি।

এই অধ্যায়ে আমরা উক্ত ভবিষ্যবাণীর সত্যতা যাচাই করার চেষ্টা করিব। পরীক্ষিত তথ্য এবং সম্ভাব্য ধারণার মধ্যে কিছু কিছু পার্থক্য দেখা যাইতে পারে। এই পার্থক্য কি দৈবাৎ হইয়াছে—না—উক্ত দুইটি বিষয়ে নিশ্চিত পার্থক্য বিদ্যমান—এই সম্বন্ধে কোন সিদ্ধান্ত করিবার পূর্বে আমাদিগকে বিশেষ বিশেষ পরীক্ষার দ্বারস্থ হইতে হইবে। Null hypothesis সেই রকম একটি পরীক্ষা। Null hypothesis আমাদিগকে জানায় যে পরীক্ষিত তথ্য এবং সম্ভাব্য ধারণার মধ্যে নির্ভর যোগ্য পার্থক্য নাই। পার্থক্য যদিও থাকে, তবে তাহা দৈবাৎ হইয়াছে বুঝিতে হইবে। সম্ভাবনার নিয়মানুসারে ঐ পার্থক্যের ব্যাখ্যা করা সম্ভবপর। যদি উক্ত গণ্ডীর বাইরে পড়ে তবে বুঝিতে হইবে—ঐ পার্থক্য নির্ভরযোগ্য—সম্ভাবনার নিয়মানুসারে তাহার ব্যাখ্যা করা সম্ভবপর নয়। এই অর্থে Null hypothesis আইন শাস্ত্রের সমতুল্য। আইনে বলে যে একজন লোক যতক্ষণ না দোষী বলিয়া সাব্যস্ত হইতেছে ততক্ষণ সে নির্দোষ। তাই Null hypothesis পরীক্ষার পরীক্ষা নিরূপণে সাহায্য করে। নীচে একটি উদাহরণ লওয়া হইল। কোন শিক্ষাবিদের ধারণা যে ভারত এবং আমেরিকার একই বয়স্ক ছাত্রের মধ্যে বুদ্ধির কোন পার্থক্য নাই। এখন যদি ভারত ও আমেরিকার একই বয়স্ক ছাত্রের উপর বুদ্ধির পরীক্ষা করিয়া দেখা যায় যে তাহাদের বুদ্ধির পার্থক্য বিশেষ কিছু নাই বা যাহা কিছু বিদ্যমান তাহা সম্ভাবনার নিয়মানুগ ব্যাখ্যার গণ্ডীর মধ্যে অবস্থিত—তাহা হইলে উক্ত Null hypothesis-টি অবিকৃত অবস্থায় রক্ষিত হইল। বিপরীতক্রমে Null hypothesis রক্ষিত হইবে না। শেষোক্ত ঘটনাকে এই বলিয়া ব্যাখ্যা করা

হয় যে ভারত ও আমেরিকার এক বয়সী ছাত্রদের মধ্যে বুদ্ধির পার্থক্য আছে এবং সেই পার্থক্য সম্ভাবনার অজুহাতে অবহেলা করা যাইতে পারে না।

কিন্তু Null hypothesis-টি রক্ষিত বা বিনষ্ট হইয়াছে ঘোষণা করিবার পূর্বে আমাদিগকে বিশেষ সতর্কতা অবলম্বন করিতে হইবে। যেমন নীচে Extra Sensory Perception-এর একটি উদাহরণ দেওয়া হইল। 16টি তাসের বাণ্ডেলে প্রত্যেকটি চারিটি করিয়া চারি প্রকারের তাস আছে। প্রতি তাসের ঠিক করিয়া "ডাক দিবার" সম্ভাবনা $\frac{1}{4}$. যদি কোন পরীক্ষার্থী এই সম্ভাবনা হইতে খুব বেশীবার কৃতকার্য হয়, তবে উক্ত ESP-এর Null hypothesis বিনষ্ট হইবে। কিন্তু তাই বলিয়া বিপরীত ধারণা হঠাৎ করিতে নাই। তাসগুলি যদি ভালভাবে তাসান না হয়, তাসের অবয়ব যদি বিভিন্ন হয়, চারি প্রকার প্রতিকৃতির রং যদি বিভিন্ন হয়—প্রভৃতি কারণে ঐ সম্ভাবনা অহেতুক বাড়িয়া যাইবে। সুতরাং পুনঃ পুনঃ সযত্ন-পরিচালিত পরীক্ষা-নিরীক্ষার পর Null hypothesis-কে সত্য বা মিথ্যা বলিয়া ঘোষণা করা উচিত—তাহার পূর্বে নহে।

χ^2 (chi-square)-test এবং Null hypothesis :—

Null hypothesis রক্ষিত হইবে বা বিনষ্ট হইবে—তাহা নির্ণয় করার বিভিন্ন পদ্ধতি বিদ্যমান। χ^2-test তাহাদের মধ্যে একটি। পরীক্ষিত তথ্য এবং সম্ভাব্য ধারণার মধ্যে পার্থক্যের নির্ভরযোগ্যতা যাচাই করিতে χ^2-test একটি মূল্যবান হাতিয়ার। নীচে χ^2-এর সূত্রটি প্রদত্ত হইল।

$$\chi^2 = \Sigma \left[\frac{(f_o - f_e)^2}{f_e} \right] \cdots\cdots\cdots\cdots\cdots (38)$$

যেখানে f_o=পরীক্ষিত তথ্যের সংখ্যা,

এবং f_e=সম্ভাব্য ধারণার সংখ্যা।

প্রতিবার পরীক্ষিত তথ্যের সংখ্যা এবং সম্ভাব্য ধারণার সংখ্যার অন্তর বাহির করিয়া তাহাদের বর্গ নির্ণয় করিতে হয়। ঐ বর্গের সমষ্টি হইল χ^2. উপরোক্ত সংখ্যা দুইটির মধ্যে ব্যবধান যত কম থাকিবে, χ^2 তত ক্ষুদ্র হইবে এবং পরীক্ষিত তথ্যের সংগে সম্ভাব্য ধারণার সম্বন্ধ ততই ঘনিষ্টতর হইবে। বিপরীত ক্রমে, χ^2 যত বেশী হইবে, পরীক্ষিত তথ্যের সংগে সম্ভাব্য ধারণার ব্যবধান ততই বৃদ্ধি পাইবে। χ^2 বাহির করিয়া Table E-এর সাহায্যে তাহার মূল্যায়ন করিতে হয়। মূল্যায়নে χ^2-এর মান এবং নির্দিষ্ট সংখ্যক 'degrees of freedom'-এর দরকার হয়। Degrees of freedom বা

df=(r−1) (c−1), যেখানে r=সারিতে ভর্তির সংখ্যা এবং c=স্তম্ভে ভর্তির সংখ্যা। Table E হইতে আমরা χ^2-এর সম্ভাবনা বা P বাহির করি। P, ·05 বা তাহার কম হইলে সাধারণতঃ আমরা Null hypothesis পরিত্যাগ করি। নীচে কয়েকটি উদাহরণে Null hypothesis সম্বন্ধে ধারণা দেওয়া হইতেছে।

(ক) 1 নং উদাহরণ :—"বিদ্যালয়ে পাঠ্যক্রম বহির্ভূত কার্যাবলী শিক্ষার অংগ।"—ঐ মন্তব্যে 24 জন ছাত্রের মতামত নিম্নরূপ : পক্ষে—12 জন, নিরুত্তর—6 জন এবং বিপক্ষে 6 জন। এই তথ্য উক্ত ধারণার অনুকূল বা প্রতিকূল তাহা নির্ণয় করিতে হইবে।

নীচে 22নং টেবলের উপরের সারিতে প্রাপ্ত তথ্য নির্দেশক সংখ্যাগুলিকে এবং তাহার নীচের সারিতে সমান ভাগে বিভক্ত সংখ্যাগুলিকে সাজান হইয়াছে। ঐ টেবলের নীচে যথাক্রমে f_o-f_e, $(f_o-f_e)^2$, $(f_o-f_e)^2/f_e$ সারিগুলিকে রাখা হইয়াছে এবং পরে 38 নং সূত্রানুসারে χ^2-এর মান বাহির করা হইয়াছে।

	পক্ষে	উত্তর		
		নিরুত্তর	বিপক্ষে	
প্রাপ্ত সংখ্যা=f_o	10	7	7	24
সম্ভাব্য সংখ্যা=f_e	8	8	8	24
f_e-fe	2	1	1	
$(f_o-f_e)^2$	4	1	1	
$(f_o-f_e)^2/f_e$	·5	·12	·13	

$$\chi^2=\Sigma\left[\frac{(f_o-f_e)^2}{f_e}\right]=\cdot 76$$

df=2

P=·70 [Table E]

22 নং টেবল : χ^2—test

উপরোক্ত টেবলে degrees of freedom বা df =(r−1) (c−1)= (3−1) (2−1)=2. অর্থাৎ উপরের সারির দুইটি কোষ ভর্ত্তি করিলে তৃতীয়টি স্বাভাবিক ভাবে পূর্ণ হইবে। কারণ প্রথমটি=10, দ্বিতীয়টি=7 হইলে, তৃতীয়টিকে 7 হইতেই হইবে। শেষোক্ত কোষটি পরাধীন। সুতরাং সারিতে

২টি degree of freedom রহিয়াছে। এইরূপে স্তম্ভে 1টি degree of freedom রহিয়াছে। সুতরাং ঐ টেবলে মোট $2 \times 1 = 2$টি degree of freedom রহিয়াছে।

Table E-তে দুটি degree of freedom-এর সারি বরাবর $\chi^2 = \cdot 76$-এর জন্য P-এর মান হইবে প্রায় ·70.

উক্ত ফলদ্বারা এই সিদ্ধান্তে আসিতে পারা যায় যে Null hypothesis স্বীকার করিয়া বারবার মতামত জিজ্ঞাসা করিলে শতকরা 70টি সময়ে $\chi^2 = \cdot 76$ পাওয়া যাইবে। যে হেতু এখানে $P = \cdot 70$, এখানে Null hypothesis রক্ষিত হইল। ইহার ফলে এই বুঝা গেল যে প্রাপ্ত তথ্য এবং সম্ভাব্য ধারণার মধ্যে যে ব্যবধানটুকু দেখা গিয়াছিল তাহা দৈবাৎ প্রাপ্ত হইয়াছে। 'chance'-দ্বারা তাহার ব্যাখ্যা সম্ভবপর, ঐ ব্যবধান উল্লেখযোগ্য নয়।

(খ) শিক্ষা এবং মনস্তত্ত্বের বহু ক্ষেত্রে সম্ভাব্য ধারণা সমভাবে বিভক্ত না হইয়া Normal distribution-এর নিয়মানুযায়ী বিভক্ত হয়। এরূপ ক্ষেত্রে Null hypothesis যাচাই করিবার জন্য χ^2—test বিশেষভাবে সাহায্য করে। নীচে 2নং উদাহরণের মাধ্যমে তাহাকে ব্যাখ্যা করা হইল।

2নং উদাহরণ :—32 জন ছাত্রকে বুদ্ধির ক্রমানুসারে তিনটি ভাগে বিভক্ত করা হইয়াছে। এই বিভাগ এবং Normally distributed ঐ ছাত্রগণের বুদ্ধির বিভাগের মধ্যে উল্লেখযোগ্য ব্যবধান নাই। "মেধাবী ছাত্রের সংখ্যা = 12, মাঝারী = 16 জন এবং অল্পধী = 4 জন। 23 নং টেবলে পূর্বের ন্যায় χ^2—test নির্ণয় করা হইয়াছে। দ্বিতীয় সারিটি Normal distribution-এর অনুপাতে সাজান হইয়াছে। এখানে সম্ভাব্য ধারণার সারিটি সমভাগে বিভক্ত নয়। Normal probability curve-এর ভূমিটিকে 2σ—দৈর্ঘ্যে তিনটি সমান ভাগে বিভক্ত করা হইয়াছে এবং Table A হইতে নিম্নলিখিত অনুপাতগুলি পাওয়া গেল।

	অনুপাত
$+3{\cdot}00\sigma$ হইতে $+1{\cdot}00\sigma$ পর্যন্ত	·16
$+1{\cdot}00\sigma$,, $-1{\cdot}00\sigma$,,	·68
$-1{\cdot}00\sigma$,, $-3{\cdot}00\sigma$,,	·16
	1·00

	মেধাবী	মাঝারী	অল্পধী	
প্রাপ্ত সংখ্যা = f_o	12	16	4	32
সম্ভাব্য সংখ্যা = f_e	5·12	21·76	5·12	32
$f_o - f_e$	6·88	5·76	1·12	
$(f_o - f_e)^2$	47·33	33·18	1·25	
$(f_o - f_e)^2/f_e$	9·25	1·53	·24	

$\chi^2 = 11{\cdot}02$

df= 2 $P < {\cdot}01$

23নং টেবল : Normal distribution অনুযায়ী সম্ভাব্য ধারণার χ^2—test.

এখন মোট ছাত্রসংখ্যা 32-এর আনুপাতিক মান 23নং টেবলের দ্বিতীয় সারিতে লেখা হইয়াছে। যেমন, $\cdot16 \times 32 = 5{\cdot}12$, $\cdot86 \times 32 = 21{\cdot}76$ ইত্যাদি ; উক্ত টেবলের $\chi^2 = 11{\cdot}02$ এবং $df = (3-1)(2-1) = 2$. Table E হইতে দেখা যায়—P, ·01 হইতে কম। অর্থাৎ $\chi^2 = 11{\cdot}02$ হইবার সম্ভাবনা রহিয়াছে শত ভাগের এক ভাগ। সুতরাং সম্ভাবনা নিতান্ত কম থাকায় Null hypothesis-টি বর্জিত হইল। ইহার অর্থ এই যে উক্ত ব্যবধান কেবলমাত্র দৈবক্রমে ঘটিত বলিয়া ব্যাখ্যা করা যাইতেছে না—প্রকৃত ব্যবধান রহিয়াছে।

(গ) যখন χ^2-টেবলের সারিতে লিখিত সংখ্যা বেশ বড় থাকে, তখন χ^2-test-এর মূল্য বজায় থাকে। কিন্তু সংখ্যা ক্ষুদ্র হইলে এই প্রকারে যাচাই-পদ্ধতি সব সময়ে বজায় থাকে না। সাধারণতঃ 5 হইতে কম frequency থাকিলে এই অসুবিধায় পড়িতে হয়। ইহা ছাড়াও 2×2 ধরনের টেবলে χ^2-এর নির্ণেয় মানে ভ্রান্তি থাকে। ইহাকে Yates-এর সংশোধনী পদ্ধতিতে সংশুদ্ধ করিতে হয়। নীচে উদাহরণের সাহায্যে নূতন পদ্ধতিতে χ^2-test দেখান হইল (24নং টেবল)।

পূর্বেই বলা হইয়াছে 2×2 টেবলে χ^2-test পদ্ধতি একটু ভিন্ন প্রকৃতির। 24নং টেবলে $f_o - f_e$ সারির নীচের সারিতে যতিবিহীন সংখ্যার জন্য ·5 বিয়োগ করা হইয়াছে। যেমন, 30 এবং 37 এর প্রকৃত ব্যবধান $36{\cdot}5 - 30 = 6{\cdot}5$. আবার 23 এবং 30-এর প্রকৃত ব্যবধান $30 - 23{\cdot}5 = 6{\cdot}5$. ইহাই Yates-এর সংশোধন। আবার সম্ভাবনা বা P-এর ক্ষেত্রেও ব্যতিক্রম আছে। এখানে

	সত্য	মিথ্যা	
প্রাপ্ত সংখ্যা $= f_o$	37	23	60
সম্ভাব্য সংখ্যা $= f_e$	30	30	60
$f_o - f_e$	7	7	
Yates-এর ভ্রম সংশোধন ($-\cdot5$)	6·5	6·5	
$(f_o - f_e)^2$	42·25	42·25	
$\frac{(f_o - f_e)^2}{f_e}$	1·41	1·41	

$\chi^2 = 2{\cdot}82$ $\quad P = \cdot10$

df$=1$ $\quad \frac{1}{2}P = \cdot05$ [Table E]

24নং টেবল : 2×2 টেবলে χ^2-এর মান নির্ণয়।

P-এর স্থলে $\frac{P}{2}$ বাহির করিয়া সম্ভাবনাকে যাচাই করা হইয়া থাকে। উক্ত টেবলে $\frac{1}{2}P = \cdot05$.

উক্ত 2×2 টেবলের উপরের সারিতে যে সংখ্যাগুলি রহিয়াছে তাহারা বেশ বড়। কোন কোষে 5 হইতে ক্ষুদ্রতর সংখ্যা থাকিলে উক্ত পদ্ধতি একান্তই অপরিহার্য হইবে।

Exercise—IX

1. In 108 throws of a single die, the following distribution of faces was obtained :

	Faces.						
	1	2	3	4	5	6	Total
observed frequencies :	28	23	16	8	20	13	108

Do these results constitute a refutation of the "equal probability" (Null) hypothesis ?

2. The following table represents the number of boys and the number of girls who chose each of the five possible answers to an item in an attitude scale.

	Approve strongly.	Approve	Indifferent	Disapprove	Strongly Disapprove	Total
Boys	20	25	5	20	5	75
Girls	8	13	3	13	13	50

Do these data indicate a significant sex difference in attitude toward this question ? (Verify the Null hypothesis.)

দশম অধ্যায়

অভীক্ষার স্কেল নির্ণয়

বিশেষ করিয়া Normal probability curve-এর উপর নির্ভর করিয়া বিভিন্ন পদ্ধতি শিক্ষা এবং মনস্তত্ত্বের স্কেল বাহির করিবার নিমিত্ত ব্যবহৃত হয়। এখন স্কেল কি ? কাঠিন্য বা অন্যান্য কোন গুণের ক্রমানুসারে রক্ষিত যতিবিহীন মাত্রাকে স্কেল বলে। স্কেলের একক সুবিধামত লওয়া হয়। ইহা যে পদ্ধতিতে বাহির করা হইয়াছে সেই পদ্ধতির উপর নির্ভরশীল। সাধারণতঃ স্কেলের একক সমান এবং অপরিবর্তিত থাকে।

পূর্ববর্তী অধ্যায়গুলিতে কাঁচা নম্বর লইয়া আলোচনা করা হইয়াছে। সেখানে নম্বরের স্কেলের একক অপরিবর্তিত ছিল এবং গুণ বা ক্ষমতার ক্রমবর্ধমান হারে ঐ নম্বরগুলি পরিবর্তিত হইয়াছে। কিন্তু 45 এবং 55 নম্বরের ব্যবধান কি 80 এবং 90 নম্বরের ব্যবধানের সমান ? বিভিন্ন ধাপে গুণ বা ক্ষমতার মাত্রার ব্যবধান কখনও সমান হইতে পারে না। আমাদের কাঁচা নম্বর সেখানে নিরর্থক হইয়া দাঁড়ায়।

আবার বিভিন্ন মানসিক অভীক্ষার কথা বাদ দিলেও, অধিকাংশ অভীক্ষার ফল যে সাংখ্যমানে প্রকাশিত হয় তাহার বিশেষ কোন নির্ভরযোগ্যতা নাই। অনেক সময় শতকরা হারে নম্বর দেওয়া হইয়া থাকে। শতকরা নম্বর কোন গুণের আয়তক্ষেত্রাকার বিস্তৃতির অর্থ প্রদান করে। আধুনিক সময়ে Normal probability curve-এর গুরুত্ব বিশেষভাবে বিবেচিত হইলেও, সাধারণে শতকরা নম্বরের পক্ষপাতি। এই সকল নম্বর দ্বারা কোন ব্যক্তির কোন গুণ, ক্ষমতা প্রভৃতির চরম মাপ নির্ণয় করা হয়—যদিও চরম মাপ কখনও পাওয়া যায় না।

শিক্ষা ও মনস্তত্ত্ব ব্যক্তির গুণ, আচরণ প্রভৃতি পরীক্ষা করিয়া ঘোষণা করে যে ঐগুলির চরম (Absolute) পরিমাপ করা অসম্ভব এবং অবাস্তব। তাই ব্যক্তির চরম পরিমাপ বাদ দিয়া বর্তমান মনোবিজ্ঞানীরা সমষ্টির পরিপ্রেক্ষিতে ব্যক্তির তুলনামূলক পরিমাপ করেন। সমষ্টির গড়ই হইল এই তুলনার কেন্দ্র স্বরূপ। পূর্বেই বলা হইয়াছে যে স্কেলের বিভিন্ন ধাপের সমান ব্যবধান সমান মাত্রা-বিশিষ্ট নয়। Mean ঐ তুলনার কেন্দ্র হইলেও আসল সমস্যার সমাধান ইহাতে পাওয়া যায় না। তাই নূতন ধরনের স্কেলের প্রয়োজনীয়তা অনুভূত হয়।

এ ছাড়া বিভিন্ন ধরনের স্কেলের তুলনার সময় বিশেষ অসুবিধা দেখা যায়। শিক্ষা এবং বৃত্তিতে সুপরিচালনার ক্ষেত্রে বিভিন্ন স্কেল আসিয়া পড়ে। বিভিন্ন ব্যক্তির মধ্যে এই বিভিন্নমুখী স্কেলের তুলনা কি করিয়া সম্ভব পর? তাই শিক্ষা-বিজ্ঞানীর কাছে উল্লিখিত কাঁচা নম্বরের পরিবর্তে এক মাত্রা বিশিষ্ট সাধারণ স্কেলের প্রয়োজন বিশেষভাবে অনুভূত হয়। তাঁহারা বিভিন্ন পরিমাপের Mean, বিস্তৃতি এবং শ্রেণীগুচ্ছের তুলনা নির্দেশক স্কেলের উদ্ভাবন করেন। নীচে এইরূপ তিন শ্রেণীর স্কেলের আলোচনা করা হইতেছে—(1) Standard-score scale, (2) T-scale এবং (3) Percentile-scale.

(1) Standard score scale

(a) একটি উদাহরণের সাহায্যে উল্লিখিত বিষয়টিকে বিশদভাবে আলোচনা করা হইয়াছে। কোন ছাত্র 97·5 নম্বর ইংরেজীতে, 10 নম্বর পঠনে, 19·5 নম্বর খবরে, 69·5 পঠিত বিষয়ে এবং 20·5 কার্য সম্পাদনী পরীক্ষায় পাইয়াছে। সে কি ইংরেজিতে দক্ষ এবং পঠনে কাঁচা? সে কি উক্ত সকল পরীক্ষাতে সমজাতীয় নম্বর পাইয়াছে? উক্ত কাঁচা নম্বর হইতে এরূপ কোন প্রশ্নেরই জবাব পাওয়া সম্ভব নয়। নীচে 25নং টেবলে ঐ ছাত্রের (1নং) নম্বরগুলি 4নং স্তম্ভে ক্রমান্বয়ে সাজান হইয়াছে। সকল বিষয়ে গড় নম্বরের তুলনায় ঐ ছাত্রের নম্বরের বিচার কিছুটা সম্ভব পর। Mean-গুলি 2নং স্তম্ভে লেখা হইয়াছে।

ঐ ছাত্রের নম্বরের দিকে দৃষ্টিপাত করিলে উপর উপর ধরা যায় যে সে ইংরেজী, পঠিত বিষয় ও কার্য সম্পাদনী পরীক্ষায় গড় নম্বর হইতে বেশী এবং পঠন ও সংবাদ পরীক্ষায় গড় নম্বর হইতে কম পাইয়াছে। আবার 4নং স্তম্ভে আর একটি ছাত্রের (2নং) নম্বরও দেওয়া হইয়াছে। আপাতদৃষ্টিতে এইটুকু বুঝা যায় যে সে ইংরেজী, পঠিত বিষয় এবং কার্য সম্পাদনী পরীক্ষায় প্রায় গড় নম্বর এবং পঠন ও সংবাদ পরীক্ষায় গড় নম্বর হইতে বেশী পাইয়াছে।

এখন প্রশ্ন আসিতে পারে যে কোন ছাত্র দুই বিষয়ে Mean নম্বর অপেক্ষা বেশী নম্বর পাইলে, ঐ দুই বিষয়ের কোনটিতে তাহার ব্যুৎপত্তি অপেক্ষাকৃত বেশী হইবে? উক্ত 1নং ছাত্র ইংরেজী ও কার্য সম্পাদনী অভীক্ষায় গড় নম্বর অপেক্ষা যথাক্রমে 19·7 এবং 8·1 পাইয়াছে। কোন ফল অধিকতর প্রশংসনীয়? 2নং ছাত্র mean ফল অপেক্ষা পঠন ও সংবাদে যথাক্রমে 10·1 এবং 8·7 নম্বর বেশী পাইয়াছে। সে কি উভয় বিষয়ে সমান দক্ষ?

আবার ঐ দুই ছাত্রের মধ্যে তুলনা কিরূপে সম্ভব পর? প্রথম ছাত্রটি তিন বিষয়ে এবং দ্বিতীয় ছাত্রটি অন্য দুই বিষয়ে উন্নততর দক্ষতা দেখাইয়াছে। আপাত দৃষ্টিতে এই পর্যন্ত জানা যায়। কিন্তু ধরা যাউক এই দুই ছাত্র একটি বৃত্তির জন্য প্রতিযোগিতা করিতেছে। কে সেই বৃত্তির অধিকারী হইবে? প্রথম ছাত্রের মোট নম্বর হইতেছে 217 এবং দ্বিতায় ছাত্রের মোট নম্বর 198·5 সুতরাং সাধারণ ভাবে বিচার করিলে দেখা যাইবে যে বৃত্তিটি প্রথম ছাত্রের প্রাপ্য। কিন্তু এই বিচার কি নির্ভুল?

আবার যদি আমরা এই দুই ছাত্রের বিভিন্ন বিষয়ের ফলের মধ্যে সামঞ্জস্য কতটা আছে দেখিতে চাই, তখন কেবলমাত্র mean-কে তুলনা কেন্দ্র করিলে হইবে না। তখন এমন একটি সাধারণীকৃত স্কেলের প্রয়োজন যাহার মাধ্যমে ঐ বিভিন্ন গুণাবলী-নির্দেশক পরিমাপকে এক সাধারণ মঞ্চের উপর দাঁড় করাইয়া তাহাদের তুলনা-মূলক বিচার করা সম্ভব হইবে। Standard-scores এই রকম একটি অতি প্রয়োজনীয় স্কেল প্রদান করে।

1 পরীক্ষা	2 Mean	3 σ	4 নম্বর		5 x Deniation		6 z Standard Scores	
			1 নং ছাত্র I	2 নং ছাত্র II	I	II	I	II
ইংরাজী	77·8	13·2	97·5	81	+19·7	+3·2	+1·49	+0·24
পঠন	16·9	4·1	10·0	27	−6·8	+10·1	−1·67	+2·48
সংবাদ	27·3	5·6	19·5	36	−7·7	+8·7	−1·67	+1·88
পঠিতবিষয়	43·5	12·9	69·5	42	+26·0	−1·5	+2·01	−0·12
কার্যসম্পাদনী	12·4	3·4	20·5	12·5	+8·1	+·1	+2·38	+0·03
যোগফল			217	198·5			+2·54	+4·51
Means							+0·51	+0·90

25 নং টেবল : কাঁচা নম্বরকে standard score-এ পরিবর্তন পদ্ধতি।

Standard-score scale-এর Mean শূন্য এবং Standard Deviation= 1·00 হইবে। 1σ, ·1σ, বা σ-এর কোন ভগ্নাংশকে এই স্কেলের একক হিসাবে ধরা হইয়া থাকে। উপরে 25নং টেবলে কাঁচা নম্বরকে Standard স্কেলে লইয়া যাওয়ার পদ্ধতি দেখান হইয়াছে।

কাঁচা নম্বরকে Standard-score বা z-এ পরিবর্তিত করিতে হইলে নিম্নলিখিত সূত্র ব্যবহার করিতে হয়।

$$z = \frac{X - M}{\sigma} = \frac{x}{\sigma} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(39)$$

(b) এখন 25নং টেবলে দুইটি ছাত্রের Standard Score বা z দ্বারা তুলনা করা যাইতে পারে। কিন্তু তুলনার প্রাক্কালে দুইটি শর্ত পালন করা অবশ্য কর্তব্য। প্রথমতঃ যে শ্রেণীর নম্বর লওয়া হইয়াছে—সেই শ্রেণীটি যেন সমগ্র পরীক্ষার্থীর (Population) প্রতিভূ হয়। দ্বিতীয়তঃ বিভিন্ন ক্ষমতা নির্দেশক নম্বরের বিভিন্ন বিস্তৃতি এবং শ্রেণীগুচ্ছের ধরন তাহাদের kurtosis বা skewness যেন সদৃশ হয়। এই শর্তে সীমিত গণ্ডীর মধ্যে আবদ্ধ হইয়া আমরা তুলনা মূলক আলোচনা করিব—ইহা যেন কখনও বিস্মৃত না হই।

Standard score-এর মাধ্যমে আমরা উল্লিখিত সমস্যা সমূহের সমাধান করিতে পারি। প্রথমে 1নং ছাত্রের কথা লওয়া হইল। (Mean হইতে) বিস্তৃতি হিসাবে যে পঠিত বিষয়ে প্রথম, ইংরেজীতে দ্বিতীয় এবং কার্য সম্পাদনী পরীক্ষায় সে তৃতীয় হইয়াছে। কিন্তু অপেক্ষাকৃত সংশুদ্ধ খবর আমরা তাহার z-score-এ পাই। সেখানে কার্য সম্পাদনী পরীক্ষায় সে প্রথম হইয়াছে, দ্বিতীয় হইয়াছে পঠিত বিষয়ে এবং তৃতীয় হইয়াছে ইংরেজীতে এ ছাড়া পঠন এবং সংবাদের পরীক্ষায় সে সমভাবে কম নম্বর পাইয়াছে। দ্বিতীয় ছাত্রের তথ্যও এইরূপে পাওয়া যায়।

এক্ষণে Standard-score-এর মাধ্যমে দুইটি ছাত্রের ফলের তুলনা করা হইবে। পূর্বেই বলা হইয়াছে যে কাঁচানম্বরের হিসাবে প্রথম ছাত্র পঠিত বিষয়ে অপেক্ষাকৃত বেশী নম্বর পাওয়ার সুযোগ পাইয়াছে। বিস্তৃতি (x) হিসাবেও অনুরূপ ফল পাওয়া যায়। কিন্তু Standard score-এর হিসাবে কার্যসম্পাদনী পরীক্ষায় সে এই সুযোগ পাইয়াছে। দ্বিতীয় ছাত্র পঠনে Standard-score হিসাবে বেশী সুযোগ পাইয়াছে—যদিও কাঁচা নম্বর এবং বিস্তৃতির ক্ষেত্রে ভিন্ন চিত্র পাওয়া যায়। আবার মোট কাঁচা নম্বর হিসাবে প্রথম ছাত্র দ্বিতীয় ছাত্র অপেক্ষা 7 নম্বর বেশী পাইয়াছে; কিন্তু z-score হিসাবে দেখা যায় যে দ্বিতীয় ছাত্র প্রথম ছাত্র অপেক্ষা প্রায় $0{\cdot}39\sigma$ বেশী পাইয়াছে।

সুতরাং উক্ত পরীক্ষায় দ্বিতীয় ছাত্র নিঃসংশয়ে প্রথম ছাত্র অপেক্ষা ভাল ফল করিয়াছে। কারণ দুই একটি বিষয়ে প্রথম ছাত্র অপেক্ষাকৃত বেশী নম্বর

পাইলেও তাহার ফলের মধ্যে সামঞ্জস্য নাই। অপর পক্ষে দ্বিতীয় ছাত্র সামঞ্জস্যভাবে ফল লাভ করিয়া প্রথম ছাত্রকে হারাইয়া দিল। তাই দ্বিতায় ছাত্রই বৃত্তির অধিকারী হওয়া বাঞ্ছনীয়।

Standard-score-scale-এর অসুবিধা

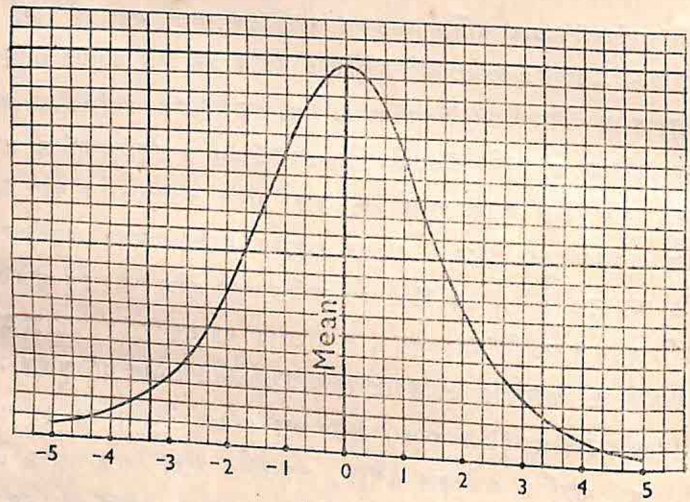
উল্লিখিত সুবিধা সত্ত্বেও Z-score-এর কিছু অসুবিধা আছে। প্রথমতঃ প্রায় অর্ধেক নম্বর ঋণাত্মক হওয়ায় হিসাবের অসুবিধা। দ্বিতীয়তঃ এককগুলি বড় দীর্ঘ—একটির পরিমাণ 1σ. আমরা Standard-score-কে একটি ধ্রুবক (মনে করি 10) দ্বারা গুণ করিয়া তাহার সহিত সুবিধাজনক অন্য একটি ধ্রুবক যোগ করিলে এই অসুবিধা দূর হইতে পারে। Standard-score-কে 10 দ্বারা গুণ করিলে একক ক্ষুদ্র হইবে—কিন্তু সমগ্র এককের সীমা বিস্তৃত হইবে। আমরা যদি Mean=50 এবং $\sigma=10$ ধরি, তাহা হইলে standardscore-কে 10 দ্বারা গুণ করিয়া 50 যোগ করিলে Standard-scale-এ নূতন নম্বর পাওয়া যাইবে।

(2) T-scale

Standard-score বাহির করিবার পদ্ধতিতে বলা হইয়াছে যে, যে সকল কাঁচা নম্বরের z-score বাহির করিতে হইবে সেই কাঁচা নম্বরের শ্রেণীগুচ্ছ পৃথিবীর সমগ্র পরীক্ষার্থীর (Population) প্রতিভূ হইবে। অর্থাৎ উক্ত শ্রেণী-গুচ্ছের Mean, σ প্রভৃতি Population-এর Mean, σ প্রভৃতির সহিত সমান হইবে। যদি উক্ত শর্তই পূরণ হয়, তাহা হইলে z-score-এর ন্যায় কাঁচা নম্বরের এককও সদৃশ হইবে। সুতরাং এককের সমতা রক্ষণের জন্য উন্নততর পদ্ধতি অনুসরণ করিয়া যে Z-score বাহির করা হইল তাহাতে অধিকতর লাভ কি হইল ? বা কি অধিকতর লাভ পাওয়াই বা সম্ভবপর ? আমরা জানি যে সাধারণতঃ আমাদের প্রাপ্ত শ্রেণীগুচ্ছ, Population-এর প্রতিভূ হয় না। কেবলমাত্র বাছাই-এর (Sample) ত্রুটির জন্য যে এই বৈষম্য দৃষ্ট হয় তাহা নহে। এই পার্থক্য অভীক্ষার কাঠিন্য, প্রশ্নগুলির মধ্যে আত্ম-অনুবন্ধ এবং কাঠিন্য ও আত্ম-অনুবন্ধের বৈষম্যের জন্যও হইতে পারে। তাই কাঁচা নম্বরকে z-score-এ রূপান্তরিত করিয়া উন্নততর চিত্র পাইলেও এককের সমতার প্রশ্নে কিছু ত্রুটি থাকিয়া যাইতেছে। Standard-score-এর এই অসুবিধা দূরীকরণের জন্য অন্য আর একটি স্কেল ব্যবহার হয়—তাহার নাম T-scale.

T-scale বাহির করিবার পদ্ধতিটি William A. Macall আবিষ্কার করেন এবং 12 বৎসর বয়স্ক ছাত্র-ছাত্রীর পঠন অভীক্ষার ফলে প্রয়োগ করেন।

(a) Standard-score-কে Normal curve-এ লইয়া যাইলে T-score পাওয়া যায়। T-scale-এর Mean=50 এবং $\sigma=10$. Standard score বা σ=score-এর Mean=0 এবং $\sigma=1$. এখন যদি Mean-কে Normal curve-এর Mean হইতে বামে 5σ একক দূরে সরাইয়া লওয়া যায়, তাহা হইলে স্কেলটি -5σ হইতে $+5\sigma$-এর পরিবর্তে 0σ হইতে 10σ পর্যন্ত বিস্তৃত হইবে। এই স্কেলে এখন Mean=50 পাওয়ার জন্য σ-কে 10 দ্বারা গুণ করা হইল। তাহা হইলে অন্যান্য বিভাগগুলি বাম দিক হইতে যথাক্রমে 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 এবং 100 হইবে। σ-score এবং T-score-এর তফাৎ 29নং চিত্রে দেখান হইল। মনে রাখা দরকার যে T-scale-এর সীমা 0 হইতে 100 পর্যন্ত ; ইহার একক বা T=1 এবং Mean=50. যেহেতু $\sigma=10$, সুতরাং $T=1\sigma$. T-scale-এর নিম্ন সীমা -5σ এবং উচ্চ সীমা $+5\sigma$ হইলেও অধিকাংশ ক্ষেত্রে T-score-গুলি $-3{\cdot}5\sigma$ হইতে $+3{\cdot}5\sigma$ অর্থাৎ 15 হইতে 85 পর্যন্ত বিস্তৃত থাকে।



Mean

[σ-scale (Mean হইতে আরম্ভ)]

0 1 2 3 4 5 6 7 8 9 10

[σ-scale (-5σ হইতে আরম্ভ)]

0 10 20 30 40 50 60 70 80 90 100

[T-scale (-5σ হইতে আরম্ভ)]

29 নং চিত্র : σ-scale-কে T-scale-এ রূপান্তরিত করা হইয়াছে।

T-scale নির্ণয় পদ্ধতি

নীচে একটি উদাহরণের সাহায্যে কাঁচা নম্বরকে T-scale-এ লইয়া যাওয়া হইয়াছে। ঐ পদ্ধতির বিভিন্ন ধাপগুলি নীচে প্রদত্ত হইল।

(i) অধিক সংখ্যক প্রতিভূ প্রশ্ন লইয়া একটি প্রশ্ন গুচ্ছ তৈরী করা হইল। আন্দাজ মত কাঠিন্যের ক্রমানুসারে তাহাদিগকে সাজান হইল। এই গুচ্ছটিকে যে বিশেষ শ্রেণীর জন্য তৈরী করা হইতেছে সেই শ্রেণীর প্রতিভূ দলের উপর প্রয়োগ করা হইল।

(ii) প্রতি প্রশ্ন শতকরা কত জন ছাত্র উত্তর করিয়াছে তাহার উপর ভিত্তি করিয়া পুনরায় তাহাদিগকে কাঠিন্যের ক্রমানুসারে সাজান হইল।

(iii) এখন ঐ নূতন গুচ্ছটি প্রতিভূ দলের উপর প্রয়োগ করা হইল এবং মোট অভীক্ষার ফল লিখিত হইল। এই মোট ফলকে 26 নং টেবলে প্রদত্ত পদ্ধতিতে T-scale-এ লইয়া যাওয়া হইল।

(1) অভীক্ষার নম্বর	(2) f.	(3) cum. f	(4) নম্বরের নীচের cum.f. + প্রদত্ত নম্বরের f-এর অর্ধেক	(5) 4নং স্তম্ভের শতকরা হিসাব	(6) T-score
10	2	50	49·0	98·0	71
9	8	48	44·0	88·0	62
8	18	40	31·0	62·0	53
6	5	22	19·5	39·0	47
5	14	17	10·0	20-0	42
3	3	3	1·5	3·0	31

26 নং টেবল : কাঁচা নম্বরকে T-score-এ পরিবর্তন পদ্ধতি।

26 নং টেবলের 1নং স্তম্ভে অভীক্ষার নম্বর এবং 2 নং স্তম্ভে উক্ত নম্বর প্রাপ্ত পরীক্ষার্থীর সংখ্যা লিখিত হইয়াছে। 3 নং স্তম্ভে নীচ হইতে cum.f বাহির করা হইয়াছে। 4 নং স্তম্ভে নম্বরের নীচের cum.f এবং প্রদত্ত নম্বরের f-এর অর্ধেক লওয়া হইয়াছে। যেমন—সর্বনিম্ন ধাপে $0+\frac{1}{2}\times 3=1{\cdot}5$, তাহার উপরের ধাপে $3+\frac{1}{2}\times 14=10{\cdot}0$, তাহার উপরের ধাপে $17+\frac{1}{2}\times 5=19{\cdot}5$

ইত্যাদি। 5নং স্তম্ভে 4নং স্তম্ভের শতকরা হারের হিসাব লেখা হইয়াছে। পরে G টেবল হইতে ঐ শতকরা নম্বর বরাবর Normal curve-এর নম্বর বাহির করিয়া 6নং স্তম্ভে লেখা হইয়াছে।

পূর্বে T-scale-এর উপকারিতা সম্বন্ধে বলা হইয়াছে। উপরে T-score তৈরী পদ্ধতিতে বুঝা যাইতেছে যে কাঁচা নম্বরের percentile rank-ই হইতেছে T-scale. কোন অভীক্ষায় রামের PR যদি 69 হয়, তবে তাহার T-score হইবে 55 (G টেবল)। অর্থাৎ 55 নম্বরের নীচে Normal curve-এর 69% ক্ষেত্র আছে। বর্তমানে শিক্ষা এবং মনস্তত্ত্বে T-scale-এর ব্যাপক প্রয়োগ দেখা যায়। বিভিন্ন ধরনের অভীক্ষাকে T-scale-এর মাধ্যমে তুলনা করা হয়। Normal probability curve-এর উপর নির্ভর করিয়া 100-এককের মধ্যে এই স্কেলটি থাকে বলিয়া ইহার উপযোগিতা বহুল পরিমাণে দেখা যায়। কোন শ্রেণীগুচ্ছ Normal হউক বা না হউক—T-scale-এর কিছু আসে যায় না। T-scale সকল প্রকার শ্রেণীগুচ্ছকে Normal curve-এ লইয়া যায়। সেই হিসাবে σ-score-এর তুলনায় T-scale-এর দান অনেক বেশী।

(3) Percentile scale

কোন অভীক্ষায় প্রাপ্ত কোন ছাত্রের নম্বরকে percentile rank (PR)-এ লইয়া যাওয়া যায়। PR বাহির করার পদ্ধতি পঞ্চম অধ্যায়ে আলোচিত হইয়াছে। PR, 100 জন ছাত্রের স্কেলে কোন ছাত্রের অবস্থান নির্দেশ করে এবং সংগে সংগে সেই নম্বরের নীচে শতকরা কত ছাত্র অবস্থান করিতেছে তাহারও হিসাব দেয়। PR বিভিন্ন অভীক্ষার আনুপাতিক মান নির্ণয়ে সক্ষম। যে হেতু PR-কে সহজে অংক কষিয়া বাহির করা যায় এবং PR সহজে সাধারণের বোধগম্য হয়, তাহার scale-টির প্রচলন খুব বেশী হওয়া স্বাভাবিক।

কিন্তু উপরোক্ত সুবিধা সত্ত্বেও PR-scale-এ যথেষ্ট অসুবিধা আছে। PR-scale-টিকে গ্রহণ করার অর্থ হইতেছে—কোন গুণ বা ক্ষমতার আয়ত-ক্ষেত্রাকার বিস্তৃতিকে স্বীকার করিয়া লওয়া। কারণ 10 এবং 20 PR-এর ব্যবধানটি 50 এবং 60 PR-এর ব্যবধান সমান। সকল ক্ষেত্রে ইহা কিরূপে সম্ভবপর? আমরা জানি যে মানসিক এবং শিক্ষা সংক্রান্ত বহু গুণের বিস্তৃতি Normal probability curve-কে অনুসরণ করে। সুতরাং কেবল মাত্র বিশেষ বিশেষ ক্ষেত্রেই Percentile-scale-টি ব্যবহৃত হয়।

Exercise—X

1. (a) Determine the standard scores for the two students in the data given below.

(b) Give a rank order to each student is the fine tests, first in terms of raw scores, then in terms of standard scores. Explain discrepancies in rank order.

Test	1	II	III	IV	V
Mean	25	17	27	31	23
σ	3·8	5·6	7·25	4·7	6·5
Student A	30	28	32	19	37
Student B	17	31	16	34	43

2. Firee problems are passed by 13%, 36%, 50%, 68%, and 92%, respectively, of a large unselected group. If the zero point of the ability in this test is taken to be at—3σ what is the σ value of each problem as measured from this point ?

3. (a) Locate the deciles in a normal distribution. Compare the distances in terms of σ between successive 10 percent points. Explain why these distances are unequal.

(b) Divide the base line of the normal probability curve (take as 6σ) into 15 equal parts, and erect a perpendicular at each point of division. Compute the percentage of total area comprised by each division. Are these percents of area equal ? If not, explain why.

4. Find out the T scores of the given data.

Test score	f	f
12	1	4
10	3	28
9	9	13
8	7	5
6	18	16
5	20	4
3	2	1
	N=60	N=71

5. Calculate T scores for the midpoints of the class intervals in the following distribution :

Scores	f	Percent below given interval plus one-half reaching midpoint.	T score.
80	7		
75	13		
70	28		
65	16		
60	12		
55	12		
50	6		
45	1		
	·95		

———

একাদশ অধ্যায়

অনুবন্ধ-সহগ নির্ণয় করার অন্যান্য পদ্ধতি

সপ্তম অধ্যায়ে দুইটি variable-এর অনুবন্ধ-সহগ বাহির করার পদ্ধতি আলোচিত হইয়াছে। অভীক্ষার নম্বরগুলি এক যতিবিহীন স্কেলে অবস্থিত থাকে। কিন্তু শিক্ষা বা মনস্তত্ত্বে এমন মান আছে যাহাকে সংখ্যা দ্বারা প্রকাশ করা সম্ভব নয়। ব্যক্তিত্ব, মানসিক অনেক গুণ এই শ্রেণীর অন্তর্গত। তাহাদিগকে পদের ক্রমানুসারে সাজান যাইতে পারে। ইহা ছাড়া পরিসংখ্যানকে এমন সবগুণের সম্মুখীন হইতে হয় যে সেখানে Product moment r একেবারেই অচল। সেই সকল স্থানে কি ভাবে অনুবন্ধ-সহগ নির্ণয় করা যায় নীচে তাহার পদ্ধতি দেওয়া হইল।

(A) পদমর্যাদা (ranks) হইতে অনুবাদ নির্ণয়

একটি গুণ বিভিন্ন ছাত্রের বিভিন্ন মাত্রায় বিদ্যমান। ছাত্রদিগকে মানের ক্রমানুসারে 1, 2, 3 প্রভৃতি পদ মর্যাদায় অভিষিক্ত করা যাইতে পারে। হাতের লেখা ছবির গুণ প্রভৃতি ক্ষেত্রে এই পদমর্যাদা প্রদান করিয়া সাজান যাইতে পারে। তাই যখন এইরূপ গুণ সমূহের অনুবন্ধ বা সময় নষ্ট না করিয়া তাড়াতাড়ি অল্পায়াসে অনুবন্ধ বাহির করার দরকার হয় তখন আমরা এইরূপ পদমর্যাদায় ক্রমান্বয়ে সাজাইয়া Rank-Difference পদ্ধতির সাহায্যে অনুবন্ধ সহগ ৪ (রো) বাহির করিতে পারি। মনে রাখা দরকার যে N, 25 এর কম হইলে এই পদ্ধতি অবলম্বন করা যাইতে পারে, নতুবা নহে।

নীচে f বাহির করিবার সূত্র ও পদ্ধতি দেওয়া হইল।

$$r \text{ বা } \rho = 1 - \frac{6 \times \Sigma D^2}{N(N^2-1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots(40)$$

কোন প্রদেশে A, B, C, D, E, F, G—এই সাতটি পত্রিকার জনপ্রিয়তাকে দুইটি বিচারকের (X এবং Y) কাছে দেওয়া হইল। তাঁহারা এই

সাতটি কাগজকে পদ মর্যাদা দিয়া পাঠাইলেন। এখন ঐ দুই বিচারকের বিচারের অনুবন্ধ সহগ বাহির করা হইয়াছে (27নং টেবল)।

পত্রিকা	বিচারক (X)	বিচারক (Y)	D		D^2
(1)	(2)	(3)	স্তম্ভ 2	স্তম্ভ 3	
			+	—	
A	1	2	...	1	1
B	2	1	1	...	1
C	3	6	...	3	9
D	4	3	1	...	1
E	7	4	3	...	9
F	6	5	1	...	1
G	5	7	...	2	4
			6	6	26

$$\rho = 1 - \frac{6 \times 26}{7 \times 48} = \cdot 54 \qquad \ldots\ldots(40)$$

27 নং টেবল : Rank-Difference-এর মাধ্যমে অনুবন্ধ সহগ নির্ণয়।

'ρ' বাহির করার পদ্ধতি : 1নং স্তম্ভে পত্রিকা নির্দেশক অক্ষরগুলিকে লেখা হইয়াছে। 2নং স্তম্ভে বিচারক (×) প্রদত্ত পদমর্যাদা বা rank লেখা হইয়াছে (27নং টেবল)। 3নং স্তম্ভে বিচারক Y প্রদত্ত rank লেখা হইয়াছে। পরবর্তী D স্তম্ভকে দুইটি ভাগ করিয়া (+ এবং −) 2নং এবং 3নং স্তম্ভের বিয়োগফল লেখা হইয়াছে। ফল ধনাত্মক হইলে '+' এবং ঋণাত্মক হইলে '−' স্তম্ভে লেখা হয়। এই রূপে D স্তম্ভটি পূর্ণ হইলে লক্ষ্য করা যায় যে '+' এবং '−' স্তম্ভের যোগফল সমান। পরে 3নং স্তম্ভটির বর্গ নির্ণয় করিয়া D^2 স্তম্ভটি পূর্ণ করিতে হয়। এবং 38 নং সূত্রের সাহায্যে অনুবন্ধ সহগ 'ρ' বাহির করা হয়। ρ-এর আর এক নাম Spearman-এর rank correlation coefficient এখানে 'ρ'-এর মান হইয়াছে ·54.

কিন্তু যখন একই পদমর্যাদা বা নম্বর দুই বা ততোধিক বার আসে তখন একটু ভিন্ন নিয়ম অনুসরণ করিতে হয়। নীচে 28নং টেবলে তাহা দেখান হইয়াছে।

1	2	3	4	5	6	7
ছাত্র	বিষয় (X)	বিষয় (X)	X-এর rank	Y-এর rank	D	D^2
মিহির	65	48	6	6·5	−·5	·25
আনন্দ	80	56	1	2·5	−1·5	2·25
নারায়ণ	65	56	6	2·5	3·5	12·25
দিলিপ	79	60	2	1	1	1
অনিল	72	46	3·5	8	−4·5	20·25
জগদীশ	65	51	6	4	2	4
বিজয়	72	50	3·5	5	−1·5	2·25
রূপেন	60	48	8	6·5	1·5	2·25
					0	44·50

$$ϱ = 1 - \frac{6 \times 44 \cdot 50}{8 \times 63} = \cdot 47 \quad \ldots\ldots(40)$$

28নং টেবল : দুই বা ততোধিক একই নম্বরের ϱ নির্ণয়।

28নং টেবলে ϱ বাহির করার পদ্ধতি 27নং টেবলেরই অনুরূপ ; তবে একই নম্বর দুই বা ততোধিক বার আসিলে তাহার rank কিরূপে বাহির করিতে হয় তাহা দেখান হইয়াছে। X বিষয়ের উচ্চতম নম্বর 80. সুতরাং তাহার rank '1' হইবে। পরবর্তী নম্বর 79, তাহার rank হইবে 2. তাহার পরবর্তী পদ 72 দুইবার আসিয়াছে। সুতরাং 3 এবং 4 rank-এ 72 থাকার জন্য তাহার rank হইবে 3 এবং 4 এর মাঝামাঝি অর্থাৎ 3·5. এখন 4 rank-টি নষ্ট হইল। এখন পরবর্তী পদ 65. ইহা তিনবার আসিয়াছে, অর্থাৎ 5, 6, 7 rank তিনটির আসনে আছে। সুতরাং ঐ তিনটি rank-এর মাঝামাঝি 6 হইবে ঐ তিনটি নম্বরের rank. এইরূপে 4 এবং 5 স্তম্ভ দুইটি পূরণ করা হইয়াছে। পরে পূর্ববৎ 6 এবং 7 স্তম্ভ দুইটি পূর্ণ করিয়া 40নং সূত্রানুযায়ী ϱ বাহির করা হইয়াছে। এখানে 'ϱ'-এর মান হইয়াছে ·47.

পূর্বেই বলা হইয়াছে যে 25 জন ছাত্রের কম থাকিলে বা সরল রৈখিক নয় এমন সব গুণাবলীর অনুবন্ধ সহগ 'ϱ'-এর পদ্ধতিতে বাহির করা হয়। এটা ঠিক যে ϱ ঠিক প্রকৃত অনুবন্ধ-সহগ নয়। তবে তাড়াতাড়ি ও অল্পায়াসে সহগ বাহির করা হয় বলিয়া পরিসংখ্যানবিদ্‌গণ প্রাথমিক ভাবে 'ϱ' বাহির

করেন এবং যদি দেখেন যে নির্ভরযোগ্য সহগ সেখানে বিদ্যমান, তখন r-এর শরণাপন্ন হন। সেইজন্য প্রাথমিক পরীক্ষা হিসাবে কম সংখ্যক ছাত্রের ক্ষেত্রে '৫'-এর ব্যবহার আছে।

মনে রাখা দরকার যে '৫'-এর SE বাহির করার দরকার অনুভূত হয় না।

B. Biserial correlation নির্ণয়

পূর্বে একটি variable-এর সহিত অন্য একটি variable-এর অনুবন্ধ-সহগ বাহির করা হইয়াছে। এমন কতকগুলি গুণ বা বৈশিষ্ট্যের মধ্যে ঐ সহগ বাহির করিতে হইতে পারে যাহাদের দ্বিতীয় variable-টি একক নহে—তাহার মধ্যে দুইটি ভাগ বিদ্যমান। শেষোক্ত variable-কে দ্বি-দল variable বলা হয় (dichotomous), যেমন পরীক্ষার নম্বরের সহিত খেলার ক্ষমতার তুলনা করিতে শেষোক্ত variable-কে খেলায় সক্ষম ও খেলায় অক্ষম—এই দুই দলে ভাগ করা যায়। এই দ্বি-দল বিভাগে সামাজিক—অসামাজিক, সিনেমা যাওয়ায় অভ্যস্ত—অনভ্যস্ত, শিক্ষিত—অশিক্ষিত প্রভৃতি দল বিদ্যমান। তাই কোন গুণ যখন জনসাধারণের মধ্যে Normally বিস্তৃত দেখা যায় তখন সেই গুণকে দ্বি-দলে বিভক্ত করিয়া আমরা biserial r বাহির করিতে পারি। Biserial r, Product-moment r-এরই অন্য একটি পরিমাপ।

(a) Biserial r নির্ণয় পদ্ধতি

29নং টেবলে অংকের নম্বরের সংগে বক্তৃতা করার ক্ষমতার r_{bis} বাহির করা হইয়াছে। মোট 150 জন ছাত্রের ফল লওয়া হইয়াছে। শ্রেণী-বদ্ধ অংকের নম্বর বরাবর বক্তৃতায় সক্ষম ও অক্ষম দল দুইটির 'f', 1নং এবং ২নং স্তম্ভে লেখা হইয়াছে এবং তাহাদের যোগফল 3নং স্তম্ভে রাখা হইয়াছে।

অংকের নম্বর	বক্তৃতায় সক্ষম দল (1)	বক্তৃতায় অক্ষম দল (2)	মোট সংখ্যা (3)
55-59	6	8	14
50-54	3	13	16
45-49	2	36	38
40-44	5	25	30
35-39	4	16	20
30-34	1	9	10
25-29	1	16	17
20-24	0	5	5
	$N_1=22$	$N_2=128$	$N=150$

27নং টেবল ঃ অংক ও বক্তৃতার biserial r.

উপরের টেবলে $M_T = 42{\cdot}03$, 150 জন ছাত্রের Mean নম্বর

$\sigma T = 10{\cdot}09$, 150 জন ছাত্রের S D.

$M_p = 45{\cdot}86$, বক্তৃতায় সক্ষম দলের Mean নম্বর

$M_q = 41.37$, বক্তৃতায় অক্ষম দলের Mean নম্বর

$p = {\cdot}147$ বক্তৃতায় সক্ষম দলের অনুপাত

$q = {\cdot}853$ বক্তৃতায় অক্ষম দলের অনুপাত

$u = {\cdot}23$ যে বিন্দুতে ·147 এবং ·853 Normal curve-এ বিভক্ত, সেই বিন্দুর উপর দণ্ডায়মান y-অক্ষের উচ্চতা।

M_p, M_q, σ_T p, q জানা থাকিলে 41 নং সূত্রানুযায়ী biserial r. বাহির করা যায়।

$$r_{bis} = \frac{M_p - M_q}{\sigma_T} \times \frac{pq}{u} \quad \text{...... ......(41)}$$

$$= \frac{45{\cdot}86 - 41{\cdot}37}{10{\cdot}09} \times \frac{{\cdot}147 \times {\cdot}853}{{\cdot}23}$$

$$= {\cdot}24.$$

Biserial r বাহির করিবার নিয়ম ঃ

(ক) প্রথমে বক্তৃতায় সক্ষম 22 জন ছাত্রের Mean নম্বর বা M_p এবং 128 জন ছাত্রের Mean নম্বর বা M_q বাহির করা হইয়াছে। উল্লিখিত উদাহরণে $M_p = 45{\cdot}86$ এবং $M_q = 41{\cdot}37$.

(খ) পরে সমগ্র ছাত্রের নম্বরের SD বা σ বাহির করিতে হইবে। এখানে $\sigma_T = 10{\cdot}09$.

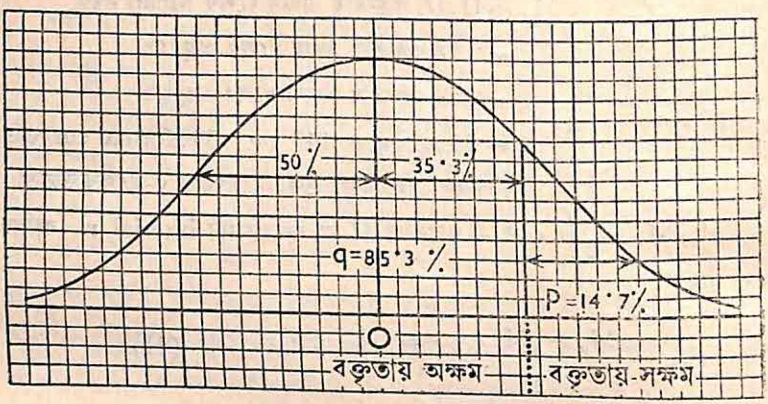
(গ) প্রথম দলে 22 জন এবং দ্বিতীয় দলে 128 জন ছাত্র আছে। অনুপাতে লইয়া যাইলে প্রথম দলের অনুপাত বা $p = {\cdot}147$ এবং দ্বিতীয় দলের অনুপাত বা $q = {\cdot}853$

বলা বাহুল্য যে সব সময় $p + q = 1{\cdot}00$ হইবে।

(ঘ) এখন বক্তৃতা করিবার গুণকে যদি Normally বিস্তৃত ধরি, তাহা হইলে একটি Normal curve আঁকিয়া (30 নং চিত্রে) আন্দাজে $p = 14{\cdot}7\%$ এবং $q = 85{\cdot}3\%$ সীমানা ঠিক করিয়া একটি ordinate টানা হইল। এখন

ঐ সীমানা দক্ষিণ অংশে Mean হইতে 35·3% দূরে অবস্থিত। Table C হইতে হিসাব করিয়া y–অক্ষের উচ্চতা পাওয়া গেল ·23.

এখন ·23 হইল দুই দলে বিভক্ত বিন্দুর উপর y—অক্ষের উচ্চতা। $\therefore$ u = ·23.



30নং চিত্র : Biserial r বাহির করিবার কালে u—এর উচ্চতা নির্ণয়।

(ঙ) এখন M_p, M_q, σ, p, q এবং u--এর মান 39নং সূত্রে স্থাপন করিলে r_{bis} পাওয়া যাইবে। এখন 29নং টেবলে r_{bis} = ·24.

Product Moment r-এর ন্যায় r_{bis}-এর SE বাহির করা দরকার। যদি N বেশ বড় হয় এবং p ও q প্রত্যেকে কম পক্ষে ·1 থাকে তাহা হইলে নীচের 42নং সূত্র দ্বারা r_{bis}-এর SE বাহির করিতে পারা যায়।

$$\sigma_{r_{bis}} = \frac{\frac{\sqrt{pq}}{u} - r^2_{bis}}{\sqrt{N}} \qquad \text{............(42)}$$

29নং টেবলে $\sigma_{r_{bis}}$ হইবে ·12

ইহা হইতে আমরা এই সিদ্ধান্তে আসিতে পারি যে 68·26% r_{bis} ·12 হইতে ·36-এর (·24 ± ·12) মধ্যে থাকিবে। নির্ভরশীলতার গণ্ডী যদি বাড়াইয়া দিই তবে বলা যাইবে যে প্রায় 95% r_{bis} ·01 হইতে ·47-এর (·24 ± 1·92 × ·12) মধ্যে অবস্থিত থাকিবে।

(b) Point Biserial r নির্ণয় :

এমন অনেক অভীক্ষা আছে যাহাদের উত্তর হইবে—সত্য-মিথ্যা, পাশ-ফেল, হাঁ-না, বা যেখানে উত্তর ঠিক হইলে '1' এবং ভুল হইলে '0' দেওয়া হয়—সেইখানে Point-biserial r, biserial r অপেক্ষা অধিক কার্যকারী। একটি variable-কে দুইভাগে বিভক্ত করিলে যতিবিহীন স্কেলে তাহারা দুইটি পৃথক ভরকেন্দ্রের সৃষ্টি করে—Point biserial এই নীতির উপর প্রতিষ্ঠিত। পুরুষ-মহিলা, জীবিত-মৃত প্রভৃতি এই শ্রেণীর অন্তর্গত। Point biserial-এর সূত্র নীচে প্রদত্ত হইল (43নং সূত্র)।

$$r_{p_{bis}} = \frac{M_p - M_q}{\sigma} \cdot \sqrt{pq} \qquad \text{............ (43)}$$

যেখানে M_p এবং M_q যথাক্রমে দুইটি দলের Mean, p এবং q দুইটি দলের অনুপাত এবং σ হইতেছে সমস্ত দলের SD.

$r_{p_{bis}}$ নির্ণয় করার পদ্ধতি 30 নং টেবলে প্রদত্ত হইল।

ছাত্রী	বুদ্ধির অভীক্ষার ফল (X)	18নং প্রশ্নের উত্তর করে নাই (Y)	X^2
রাধিকা	38	1	1444
আরতি	32	1	1024
বীণা	25	0	625
মঞ্জু	30	1	900
শান্তি	22	0	484
মৈত্রেয়ী	23	1	529
অণু	27	0	1129
রত্না	36	1	1296
তারা	35	1	1225
স্নিগ্ধা	27	0	1129
শেফালি	33	1	1089
মায়া	31	1	961
প্রণতি	29	1	841
যোগফল	388	9	12676

30নং টেবল : Point r_{bis} নির্ণয় পদ্ধতি

N_1 (পাশ সংখ্যা) = 9 $\qquad$ p = ·69

N_2 (ফেল সংখ্যা) = 4 $\qquad$ q = ·31

$$M_1 = \frac{273}{9} = 31{\cdot}89 \qquad \sigma t = 9{\cdot}17$$

$$M_2 = \frac{101}{4} = 25{\cdot}25$$

$$M_t = \frac{388}{13} = 29{\cdot}85.$$

$$r_{pb_{is}} = \frac{31{\cdot}89 - 25{\cdot}25}{9{\cdot}17} \times \sqrt{{\cdot}69 \times {\cdot}31} \qquad \ldots\ldots\ldots\ldots(43)$$

$$= {\cdot}33$$

মনে রাখা দরকার যে Product Moment r-ই Point biserial r. উক্ত 30 নং টেবলের সংখ্যাগুলিকে লইয়া Product moment পদ্ধতিতে r বাহির করিলে একই ফল পাওয়া যায়।

C. Tetrachoric r নির্ণয় :

পূর্বে একটি যতি বিহীন Variable-এর সংগে অন্য একটি দ্বি-দল variable-এর অনুবন্ধ সহগ, biserial r বা Point biserial r-দ্বারা প্রকাশ করা হইয়াছে। কিন্তু এমন variableও থাকিতে পারে যাদের প্রত্যেকের দুইটি করিয়া দল বিদ্যমান। এইরূপ (2×2) চতুর্বর্গ তালিকার অনুবন্ধ সহগকে Tetrachoric r বা r_t-দ্বারা প্রকাশ করা হয়। যখন দুই variable-এর কোনটিকেই ঠিক নম্বর দ্বারা প্রকাশ করা যাইতেছে না—অথচ প্রতিটিই দ্বি-দলসমন্বিত—এইরূপক্ষেত্রে ঐ দুই Variable এর অনুবন্ধ সহগ হইবে r_t. ঐ দুই variable যতিবিহীন এবং Normally বিস্তৃত হইবে। যেমন শিক্ষা—সামাজিক খাপ খাওয়ানো, বিবাহিত জীবন—সুখী জীবন প্রভৃতির অনুবন্ধ সহগকে r_t-দ্বারা প্রকাশ সম্ভবপর। নীচে 31 নং টেবলে r_t বাহির করার সূত্র ও পদ্ধতি দেওয়া হইল।

X. Variable (বিবাহিত মানুষ)

Y. Variable	স্বামী বা স্ত্রী মৃত	স্বামী বা স্ত্রী জীবিত	মোট
সুখী	28 (B)	45 (A)	73 (A+B)
অসুখী	34 (D)	13 (C)	47 C+D
মোট	62 (B+D)	58 (A+C)	N=120 A+B+C+D

$$r_t = \cos\left(\frac{180^\circ \times \sqrt{BC}}{\sqrt{AD}+\sqrt{BC}}\right) \quad \dots\dots\dots\dots\dots(44)$$

$$= \cos\left(\frac{180^\circ \times \sqrt{364}}{\sqrt{1530}+\sqrt{364}}\right)$$

$$= \cos 60^\circ$$

J টেবল হইতে $r_t = \cdot 50$

31নং টেবল : Tetrachoric r নির্ণয় ।

31নং টেবলে দ্বি-দল দুইটি Variable-কে সাজাইয়া রাখা হইয়াছে। X-Variable বিবাহিত মানুষের (স্বামী/স্ত্রীমৃত এবং স্বামী/স্ত্রী-জীবিত) সংগে y-variable-এর (সুখী এবং অসুখী) অনুবন্ধ সহগ বা r_t বাহির করা হইয়াছে। চতুর্বর্গ (2×2) টেবলে 120 জন লোকের নিকট হইতে প্রাপ্ত খবর অনুযায়ী A, B, C, D কোষে সাজান হইয়াছে। A+B, C+D, B+D, A+C এবং পরিশেষে A+B+C+D-এর মান টেবলে লিখিত হইয়াছে। যখন A+B এবং C+D প্রায় সমান এবং B+D এবং A+C প্রায় সমান, তখন r_t-কে 44 নং সূত্র দ্বারা বাহির করা হয়। নতুবা বীজগাণিতিক জটিল সূত্র দ্বারা r_t-নির্ণয় সম্ভব। যখন BC, AD অপেক্ষা বৃহত্তর তখন সূত্রে লবে $\sqrt{BC}$-এর পরিবর্তে $\sqrt{AD}$ লেখা হয়। r_t—cosine-এ বাহির হয় বলিয়া J টেবল হইতে তাহার মান নির্ণয় করা হয়। 31নং টেবলে $r_t = \cdot 5$. ইহা হইতে আমরা এই সিদ্ধান্তে আসিতে পারি যে জীবিত স্বামী/স্ত্রীর সহিত সুখী জীবনের সম্পর্ক বিদ্যমান ($r_t = \cdot 50$)। অবশ্য আমাদের এই সিদ্ধান্ত 120 জন লোকের মধ্যে সীমাবদ্ধ।

r_t-এর SE-নির্ণয় পদ্ধতি খুবই জটিল। তবে এইটুকু বলা যায় যে r যখন

বেশ বড় হয় এবং X এবং y-এর বিভাজন দুইটি Variable-এর Mdn-এর নিকটবর্তী হয়—তখন নিঃসন্দেহে r_t-এর যাথার্থতা ঘোষণা করা যায়।

(D) Contingency সহগ বা C নির্ণয় ঃ

দুই বা ততোধিক ভাগে বিভক্ত দুইটি Variable-এর মধ্যে অনুবন্ধ সহগের নাম Contingency সহগ বা C. 32 নং টেবলে (4×4) ছকে C বাহির করার পদ্ধতি দেখান হইয়াছে।

পিতার বুদ্ধি

পুত্রের বুদ্ধি ↑ / পিতার বুদ্ধি →	বেশী বুদ্ধিমান	বুদ্ধিমান	মাঝারি	বোকা	যোগফল
বেশী বুদ্ধিমান	(87) 125	(79) 70	(74) 61	(64) 45	301
বুদ্ধিমান	(69) 75	(62) 95	(59) 50	(51) 20	240
মাঝারি	(43) 23	(39) 36	(37) 59	(31) 31	149
বোকা	(89) 65	(79) 55	(77) 75	(65) 115	310
যোগফল	288	256	245	211	1000

32নং টেবল ঃ Contingency সহগ বা C নির্ণয়।

(a) Independence Values :

$$\frac{288\times301}{1000}=87 \qquad \frac{245\times301}{1000}=74$$

$$\frac{288\times240}{1000}=69 \qquad \frac{245\times240}{1000}=59$$

$$\frac{288\times149}{1000}=43 \qquad \frac{245\times149}{1000}=37$$

$$\frac{288\times310}{1000}=89 \qquad \frac{245\times310}{1000}=77$$

$$\frac{256\times301}{1000}=79 \qquad \frac{211\times301}{1000}=64$$

$$\frac{256\times240}{1000}=62 \qquad \frac{211\times240}{1000}=51$$

$$\frac{256\times149}{1000}=39 \qquad \frac{211\times149}{1000}=31$$

$$\frac{256\times310}{1000}=79 \qquad \frac{211\times310}{1000}=65$$

(b) C নির্ণয় :

$$\frac{(125)^2}{87} = 179{\cdot}6$$

$$\frac{(75)^2}{69} = 81{\cdot}5$$

$$\frac{(23)^2}{43} = 12{\cdot}3$$

$$\frac{(65)^2}{89} = 47{\cdot}5$$

$$\frac{(70)^2}{79} = 62{\cdot}0$$

$$\frac{(95)^2}{62} = 145{\cdot}6$$

$$\frac{(36)^2}{39} = 32{\cdot}2$$

$$\frac{(55)^2}{79} = 38{\cdot}3$$

$$\frac{(61)^2}{74} = 50{\cdot}3$$

$$\frac{(50)^2}{59} = 42{\cdot}4$$

$$\frac{(59)^2}{37} = 94{\cdot}1$$

$$\frac{(75)^2}{77} = 73{\cdot}1$$

$$\frac{(45)^2}{64} = 31{\cdot}6$$

$$\frac{(20)^2}{51} = 7{\cdot}5$$

$$\frac{(31)^2}{31} = 31{\cdot}0$$

$$\frac{(115)^2}{65} = 203{\cdot}5$$

$$S = 1132{\cdot}5$$

$$N = 1000{\cdot}0$$

$$S - N = 132{\cdot}5$$

$$C = \sqrt{\frac{S-N}{S}} \ldots\ldots\ldots\ldots\ldots(45)$$
$$= \sqrt{\frac{132\cdot5}{1132\cdot5}}$$
$$= \cdot34.$$

উক্ত টেবলে 1000 জন পিতাকে বুদ্ধির চারিটি স্তর অনুযায়ী সাজান হইয়াছে। অপর দিকে তাহাদের পুত্রদেরও অনুরূপভাবে সাজান হইয়াছে। $4 \times 4 = 16$টি কোষের মধ্যে প্রত্যেকটিকে যে যে সংখ্যায় ঐ দুই দলকে দেখা গিয়াছে তাহা রক্ষিত আছে। এ ছাড়া তাহাদের প্রত্যেকটির উপরে বন্ধনীর মধ্যে প্রতি কোষের Independence value দেওয়া আছে। Chance-এর উপর নির্ভর করিয়া বিভিন্ন বুদ্ধি নির্দেশক প্রতি সম্ভাব্য সংখ্যাই হইতেছে ঐ কোষের Independence value. যেমন উপরের সারির সর্বাপেক্ষা বামে অবস্থিত কোষের Independence value হইবে $\frac{288 \times 301}{10000}$ বা 87. তাহার দক্ষিণে অবস্থিত কোষের ঐ মান হইবে $\frac{256 \times 301}{1000}$ 79, ইত্যাদি। Independence value বাহির হইলে, প্রতি কোষের জন্য C বাহির করিতে হইবে। যেমন সব চেয়ে নীচের সারির C হইবে $\frac{65^2}{89}$ বা 47·5. ঐ সারির দক্ষিণ প্রান্তের C হইবে $\frac{115^2}{65}$ বা 203·5 ইত্যাদি। পূর্বে প্রতি কোষের জন্য C-এর মান বাহির করা হইয়াছে। ঐগুলি যোগ করিলে S পাওয়া যায়। এখন 45নং সূত্রানুযায়ী Contingency co-efficient বা C বাহির করা যায়। 32নং টেবলে C-এর মান ·34.

E. Partial এবং Multiple correlation অনুবন্ধ সহগ শিক্ষা মনোবিজ্ঞান প্রভৃতিকে নূতন দৃষ্টিভঙ্গি সহকারে দেখিতে সাহায্য করিয়াছে। ইহা দ্বারা বহু উল্লেখ যোগ্য তথ্যের রহস্যোদ্ঘাটন সম্ভব হইয়াছে। Partial এবং Multiple correlation এই তথ্যগুলির মধ্যে অন্যতম।

অনেক সময় এমন দেখা যায় যে দুইটি variable-এর অনুবন্ধ সহগ তৃতীয় একটি variable দ্বারা প্রভাবান্বিত হয়। এই অতিরিক্ত প্রভাবের ফলে হয়ত পূর্বোক্ত দুই variable-এর r অযথা বর্দ্ধিত হয়। যদি ঐ প্রভাবশালী variable-টি না থাকিত তাহা হইলে হয়ত r অনেক কম এমনকি শূন্যও হইতে

পারিত। মনে করি বুদ্ধির সহিত অর্জিত জ্ঞানের মধ্যে r বাহির করিতে হইবে। এখন বুদ্ধির সহিত বয়সের এবং অর্জিত জ্ঞানের সংগে বয়সের সম্বন্ধ অঙ্গাঙ্গিভাবে জড়িত। এখন ঐ অসুবিধা সৃষ্টিকারী প্রভাবটিকে উভয় ক্ষেত্রে একই রাখিতে না পারিলে বা বিতাড়িত করিতে (partial out) না পারিলে বুদ্ধিও অর্জিত জ্ঞানের মধ্যে প্রকৃত r বাহির করা একরূপ অসম্ভব। তৃতীয়, চতুর্থ, প্রভৃতি প্রভাব মুক্ত স্বাধীন variable দুইটির অনুবন্ধকে Partial correlation বলে।

$r_{12\cdot3}$ প্রভৃতি দ্বারা Partial correlation coefficient-কে লেখা হইয়া থাকে। এর অর্থ এই যে ইহা 1নং এবং 2নং variable-এর বিশুদ্ধ Partial correlation coefficient এবং ইহা 3নং variable-এর প্রভাব মুক্ত। $r_{12\cdot345}$ দ্বারা এই বুঝায় যে ইহা 3, 4 এবং 5নং variable-এর প্রভাবমুক্ত এমন 1নং এবং 2নং স্বাধীন variable-এর Partial correlation coefficient.

আবার অনেক সময় 2, 3, 4 প্রভৃতি variable-গুলি দ্বারা গঠিত Multiple Regression Equation হইতে লব্ধ 1নং variable-এর সম্ভাব্য (predicted) মানের অভীক্ষা-লব্ধ 1নং variable-এর অনুবন্ধ বাহির করার দরকার অনুভূত হয়। এই ধরনের অনুবন্ধকে Multiple correlation বলে। ইহাকে R-দ্বারা সূচিত করা হইয়া থাকে। অনেকে ইহাকে Multiple R বলেন। যদি $R_1(_{234})=\cdot85$ হয়, তবে বুঝিতে হইবে যে 2, 3, এবং 4 নম্বর variable-দ্বারা গঠিত Multiple Regression Equation হইতে প্রাপ্ত। 1নং variable-এর সম্ভাব্য মানের সহিত প্রাপ্ত। 1নং variable-এর মানের অনুবন্ধ সহগ হইতেছে ·85.

Multiple regression equation, Partial correlation coefficient-দ্বারা তৈরী হয় এবং ইহার যাথার্থতা R-দ্বারা যাচাই হয়, নীচে প্রথমে সূত্র এবং পরে অংকের মাধ্যমে Partial correlation এবং R-কে বুঝান হইল।

(ক) Multiple correlation-এর সমীকরণ :

(a) $\bar{x}_1 = b_{12\cdot3}x_2 + b_{13\cdot2}x_3$ (Mean হইতে দূরত্ব অনুযায়ী)

$X_1 = b_{12\cdot3}X_2 + b_{13\cdot2}X_3$ (নম্বর অনুযায়ী)

(b) $b_{12\cdot3}=r_{12\cdot3}\,\frac{\sigma_{1\cdot23}}{\sigma_{2\cdot13}}$

$b_{13\cdot2}=r_{13\cdot2}\,\frac{\sigma_{1\cdot23}}{\sigma_{3\cdot12}}$

$b_{23\cdot1}=r_{23\cdot1}\,\frac{\sigma_{2\cdot13}}{\sigma_{3\cdot12}}$

(c) $\sigma_{1\cdot23}=\sigma 1\sqrt{1-r_{12}{}^2}\,\sqrt{1-r^2{}_{13\cdot2}}$

$\sigma_{2\cdot13}=\sigma_{2\cdot31}=\sigma_2\sqrt{1-r^2{}_{23}}\,\sqrt{1-r^2{}_{12\cdot3}}$

$\sigma_{3\cdot12}=\sigma_{3\cdot21}=\sigma_3\sqrt{1-r^2{}_{23}}\,\sqrt{1-r^2{}_{13\cdot2}}$

(d) $r_{12\cdot3}=\frac{r_{12}-r_{13}\,r_{23}}{\sqrt{1-r_{13}{}^2}\,\sqrt{1-{}_{23}{}^2}}$

$r_{13\cdot2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{1-r_{12}{}^2}\,\sqrt{1-r_{23}{}^2}}$

$r_{23\cdot1}=\frac{r_{23}-r_{12}\,r_{13}}{\sqrt{1-r_{12}{}^2}\,\sqrt{1-r_{13}{}^2}}$

(e) X_1 এর SE

$\sigma(_{est}\,\chi_1)=\sigma_{1\cdot23}$

(f) Multiple correlation coefficient :

$R_{1(23)}=\sqrt{1-\frac{\sigma^2{}_{1\cdot23}}{\sigma_1{}^2}}$

(খ) **পরিমাপ :**

$N=75$

X_1 = criterion—সাহিত্যের নম্বর।

X_2 = ইতিহাসের নম্বর।

X_3 = ভূগোলের নম্বর।

$M_1=78\cdot00$	$M_2=87\cdot20$	$M_3=32\cdot80$
$\sigma_1=10\cdot21$	$\sigma_2=6\cdot02$	$\sigma_3=10\cdot35$
$r_{12}=\cdot67$	$\sigma_{13}=\cdot75$	$r_{23}=\cdot63$

(i) Partial r নির্ণয় :

$r_{12\cdot3}=\frac{\cdot67-\cdot67\times\cdot75}{\sqrt{1-(\cdot75)^2}\,\sqrt{1-(\cdot63)^2}}=\cdot39$

$r_{13\cdot2}=\frac{\cdot75-\cdot67\times\cdot75}{\sqrt{1-(\cdot67)^2}\sqrt{1-(\cdot63)^2}}=\cdot57$

$$r_{23\cdot1} = \frac{\cdot63 - \cdot67 \times \cdot63}{\sqrt{1-(\cdot67)^2}\sqrt{1-(\cdot75)^2}} = \cdot26$$

(ii) Partial σ নির্ণয় :

$\sigma_{1\cdot23} = 10\cdot21 \times \cdot742 \times \cdot82 = 6\cdot23$

$\sigma_{2\cdot13} = 6\cdot02 \times \cdot777 \times \cdot92 = 4\cdot31$

$\sigma_{3\cdot12} = 10\cdot35 \times \cdot777 \times \cdot82 = 6\cdot53$

(iii) Partial regression coefficient এবং regression equation নির্ণয় :

(a) $b_{12\cdot3} = \cdot39 \times \frac{6\cdot23}{4\cdot31} = \cdot55$

$b_{13\cdot2} = \cdot57 \times \frac{6\cdot23}{6\cdot53} = \cdot54$

(b) $\bar{X}_1 = \cdot55X_2 + \cdot54X_3 + 12$ (প্রায়)

(iv) $\sigma_{(est\ X_1)} = \sigma_{1\cdot23} = 6\cdot23$

(v) Multiple correlation coefficient :—

$$R_{1(23)} = \sqrt{1 - \frac{(6\cdot23)^2}{(10\cdot21)^2}} = \cdot79$$

Exercise—XI

1. Compute the correlation between the following two series of test scores by the rank-difference method and test its significance.

X	Y
135	100
153	88
138	108
145	94
126	102
124	114
118	109
147	85
126	84
88	87
76	100

110	84
101	116
135	110
123	108

2. Check the product-moment correlations obtained in problems 1 (a), 1 (b), 2 (a) & (b) and 3, Exercise—VII.

(3) The table below shows the distributions of scores on an achievement test earned by those students who answered 50% or more and those who answered less than 50% of the items in an arithmetic test correctly. compute r_{leis}.

Scores	50%+	54%
75 – 79	9	1
70 – 74	11	3
65 – 69	27	17
60 – 64	25	42
55 – 59	16	32
50 – 54	9	16
45 – 49	6	2
	103	113

4. Compute the Tetrachoric r for the following tables.

Correspondence of yes and no answers to two items in a psychological test.

Question 1

Question 2	No	Yes	Totals
Yes	65	133	198
No	90	72	162
Totals	155	205	360

5. Compute ϕ coefficients for the two tables given below. Test the significance of ϕ by way of x^2.

(a)

		Ques 1. No	Ques 1. Yes	
Ques 2.	Yes	15	90	105
	No	25	35	60
		40	125	165

(b)

		Ques 1. −	Ques 1. +	
Ques. 2.	+	13	29	42
	−	25	9	34
		38	38	76

6. Calculate the coefficient of contingency, c, for the two tables given below.

	English	Maths.	Hist.	Geogr.	Sc.	Totals
English	28	12	71	43	26	180
Maths.	6	63	38	37	36	180
Hist.	8	19	103	32	18	180
Geogr.	12	16	40	72	40	180
Sc.	4	49	26	39	62	180
	58	159	278	223	182	900

দ্বাদশ অধ্যায়

পরীক্ষা-লব্ধ মাপের Reliability এবং Validity নির্ণয়।

কোন অভীক্ষালব্ধ নম্বর তখনই নির্ভরযোগ্য বলিয়া গণ্য হইবে যখন সেই নম্বর নির্দিষ্ট এবং বিশ্বাসযোগ্য। নির্দিষ্ট মান ও নির্ভরশীলতা তখনই পাওয়া যায় যখন অভীক্ষা সত্যকার ক্ষমতার পরিমাপ দেয়—অন্য কোন ভ্রমজনিত সম্বন্ধ দ্বারা প্রভাবান্বিত হয় না। একই 'গজকাঠি' একই ব্যক্তিকে পরিমাপ করিলে প্রতি বারেই একই ফল দিবে—যদি ইতিমধ্যে যে ব্যক্তির উপর প্রয়োগ করা হইতেছে তাহার আভ্যন্তরীণ অবস্থা অপরিবর্তিত থাকে।

(a) Reliability-র নীতিগত সূত্র :—

Guilford বলেন, "The reliability of any set of measurements is logically defined as the proportion of their variance that is true variance." উপরের সংজ্ঞাটির ব্যাখ্যা প্রয়োজন। প্রথমেই বলা হইয়াছে যে পরিমাপের নির্ভরশীলতা। মনে রাখা দরকার যে আমরা পরীক্ষার হাতিয়ার অভীক্ষার নির্ভরশীলতা যাচাই করিতে যাইতেছি না—আমরা অভীক্ষালব্ধ ফলের নির্ভরশীলতা বাহির করিতে যাইতেছি। কারণ শিক্ষায় ও মনস্তত্ত্বে প্রযুক্ত অভীক্ষার ন্যায় যাহাদের উপর অভীক্ষা প্রয়োগ করা হইতেছে তাহাদের উপরেও নির্ভরশীলতা নির্ভর করে।

দ্বিতীয়তঃ, সংজ্ঞাতে 'true variance'-এর কথা বলা হইয়াছে। True variance বা σ^2_t কি? কোন পরিমাপের σ^2_t-এর অর্থ হইল—ঐ পরিমাপের Mean হইতে নম্বরগুলির দূরত্বের বর্গের গড়। Total variance, true variance এবং error variance-এর যোগফল। যেমন কোন পরিমাপের দুইটি অংশ থাকে—একটি সত্য পরিমাপ ও অপরটি তার ভুলের অংশ। ঠিক সেই রকম Total variance-ও দুইটি অংশে বিভক্ত—একটি True variance এবং অপরটি Error variance. উপরোক্ত তিনটি variance নীচের 46নং সূত্রানুযায়ী সম্বন্ধযুক্ত।

$$\sigma^2_t = \sigma^2_1 + \sigma^2_e \quad \cdots\cdots (46)$$

যেখানে σ^2_t = Total variance.

σ^2_{∞} = True variance.

এবং σ^2_e = Error variance.

পূর্বের সম্বন্ধ হইতে আমরা পাই যে,

$$\frac{\sigma^2_t}{\sigma^2_t} = \frac{\sigma^2_{\infty}}{\sigma^2_t} + \frac{\sigma^2_e}{\sigma^2_t} = 1{\cdot}00$$

এখন সংজ্ঞানুসারে $\frac{\sigma^2_{\infty}}{\sigma^2_t}$ হইতেছে reliability-সহগ, ইহা r_{tt}-দ্বারা সূচিত হয়।

$$\therefore\ r_{tt} = 1 - \frac{\sigma^2_e}{\sigma^2_t}$$

$= 1 - e^2$ [যেখানে Error variance এবং Total variance-এর অনুপাত $= e^2$]

নীচে 33নং টেবলের সাহায্যে উপরোক্ত সত্যটি প্রমাণিত হইয়াছে।

$$\sigma^2_t = \sigma^2_{\infty} + \sigma^2_e$$

নীচের টেবলে $(2{\cdot}19)^2 = (2{\cdot}05)^2 + ({\cdot}78)^2$

Reliability-সহগ $= r_{tt} = \frac{42{\cdot}0}{48{\cdot}1} = {\cdot}87$

অথবা $r_{tt} = 1 - \frac{6{\cdot}1}{48{\cdot}1} = {\cdot}87$

কিন্তু প্রতি পরিমাপে এই সম্বন্ধ ঠিক ঠিক বজায় নাও থাকিতে পারে। কারণ ভুলের পরিমাপ এক ছাড়া বেশীও হইতে পারে। আমরা এখানে একটি ভুলের উপস্থিতি স্বীকার করিয়া উক্ত সিদ্ধান্তে আসিয়াছি।

(b) Reliability নির্ণয় পদ্ধতি :

এখন আমরা নীতিগত আলোচনা বাদ দিয়া অভীক্ষা-লব্ধ পরিমাপের reliability নির্ণয় পদ্ধতি সম্বন্ধে আলোচনা করিব। বিভিন্ন পদ্ধতিকে সাধারণতঃ চারিটি শ্রেণীতে বিভক্ত করা যায়। অন্যান্য পদ্ধতি এই চারিটিরই বিশেষ অবস্থা

মাত্র। যথা—(ক) Test-Retest, (খ) Alternate-forms reliability, (গ) Split-Half পদ্ধতি এবং (ঘ) Rational equivalence.

	সত্য পরিমাপের অংশ	ভুলের অংশ	মোট পরিমাপ
	1·0	−·4	·6
	3·0	+·4	3·4
	4·0	−·8	3·2
	5·0	−·4	4·6
	5·0	+·4	5·4
	5·0	0	5·0
	5·0	+2·0	7·0
	6·0	−·8	5·2
	7·0	−·4	6·6
	9·0	0	9·0
	50·0	0	50·0
M	5·0	0	5·0
Σx^2	42·0	6·1	48·1
σ	2·05	·78	2·19
	σ_∞	σ_e	σ_t

33নং টেবল ঃ Reliability-সহগ নির্ণয়।

কোন একটি পদ্ধতিই গরিষ্ঠ নয়। বিভিন্ন চাহিদার পরিপ্রেক্ষিতে reliability বিভিন্ন পদ্ধতির সাহায্যে নির্ণীত হইবে।

(ক) **Test-Retest পদ্ধতি ঃ**

একই অভীক্ষা পুনর্বার একই ছাত্রের উপর প্রয়োগ করিবার পর ঐ দুই ফলের মধ্যে সঙ্গতি পরিলক্ষিত হইলে অভীক্ষাটি নির্ভরযোগ্য বলিয়া গণ্য হয়। সেইজন্য Test-Retest পদ্ধতি হইল reliability বাহির করিবার সর্বাপেক্ষা সহজ পদ্ধতি। ঐ দুই পরিমাপের মধ্যে উচ্চ অনুবন্ধ-সহগ বিদ্যমান থাকিলে অভীক্ষাটি নির্ভরযোগ্য হইবে।

এই পদ্ধতির কিছু অসুবিধাও আছে। প্রথমবারের উত্তরের কিছু কিছু স্মরণীয় অংশ দ্বিতীয়বারে সঞ্চালিত হইতে পারে। স্মৃতি দ্বিতীয়বারের উত্তরকে

প্রভাবান্বিত করিতে পারে। সুতরাং কম সময়ের ব্যবধানে অভীক্ষা প্রয়োগ করিলে ফল স্মৃতির দোষে দুষ্ট হয়। আবার সকল ছাত্রের সঞ্চালনের ক্ষমতা সমান নয়। সুতরাং দ্বিতীয়বারে সকলের ফল সমানভাবে প্রভাবান্বিত নয়। বিপরীতপক্ষে অভীক্ষার দুইবার প্রয়োগের মধ্যে সময়ের ব্যবধান বেশী হইলে অন্যান্য ভুলের মাত্রা বর্ধিত হয়। অভীক্ষাসমূহ অল্পবয়স্ক ছাত্রদের উপর প্রয়োগ করা হইতেছে। সময়ের ব্যবধানে তাহাদের শারীরিক, মানসিক প্রভৃতির বর্ধন পরিলক্ষিত হইবে। সকলের বর্ধনের হার সমান হইলে দোষ কম হইত; কিন্তু সকলের বর্ধনের হার সমান নয়। সাধারণতঃ বর্ধনে ফলবৃদ্ধি ঘটে এবং তাহা reliability-সহগের হ্রাস ঘটায়। উপযুক্ত সময়ের ব্যবধানে যদি অভীক্ষাটি প্রযুক্ত হয়, তবে reliability-সহগ, stability-সহগের রূপ ধারণ করে। যে সকল অভীক্ষা নূতনত্ববহুল, সেই সকল অভীক্ষার reliability-সহগ কমিয়া যায়। এই সকল কারণে Test-Retest পদ্ধতি ঠিক নির্ভরযোগ্য এবং ব্যবহার-সিদ্ধ পদ্ধতি নয় বলিয়া কমই ব্যবহৃত হয়।

(খ) অনেক সময় সমগোত্রীয় বা সমান্তরাল যোগ্যতা বিশিষ্ট দুইটি অভীক্ষা দ্বারা (Form A এবং Form B) আত্ম-অনুবন্ধ বাহির করিয়া অভীক্ষার নির্ভর-যোগ্যতা নির্ণয় করা হয়। এই পদ্ধতি মনস্তত্ত্ব বা শিক্ষাগত ফল নির্ণয়ে অধিকতর কার্যকরী। Form A অভীক্ষা প্রয়োগ করার বেশ কিছু সময় ব্যবধানে Form B অভীক্ষাটি প্রয়োগ করা হয়। এই সময় ব্যবধানে প্রথম অভীক্ষার ধরন সম্বন্ধে তাহার সম্পূর্ণ বিস্মৃতি ঘটে। মনে রাখা দরকার প্রথম অভীক্ষায় যে যে ধরন বা কৌশল অবলম্বন করিতে হয়, সেই ধরন বা কৌশল অবলম্বনে দ্বিতীয় অভীক্ষাটি প্রস্তুত করা কষ্ট সাপেক্ষ। আবার ধরন একই প্রকার হইলেও তাহাদের কাঠিন্যের মাত্রা সমান পর্যায়ে আনয়ন করাও কষ্টকর ও অভিজ্ঞতা সাপেক্ষ। এই বিষয়ে বিশেষ দৃষ্টি না দিলে যে জন্য এই পদ্ধতি অবলম্বিত হইয়াছে তাহা ব্যর্থ হইবে।

আবার Form A এবং Form B সমগোত্রীয় হওয়ার জন্য reliability সহগ উহার প্রভাবমুক্ত নয়। কিন্তু এই সমগোত্রীয়তা যদি সমগ্র অভীক্ষার মধ্যে ব্যাপ্ত থাকে, তাহা হইলে খুব বেশী ক্ষতি নাই। কারণ Form B-এর প্রতি item প্রভাবান্বিত হওয়ায়, দুই অভীক্ষার আপেক্ষিক অবস্থান প্রায় সমান হয় এবং তাহার জন্য reliability-সহগ অবিকৃত থাকে। উন্নত নির্ভরযোগ্য অভীক্ষা এই পদ্ধতি দ্বারা নির্ণীত হয়।

পরীক্ষা দ্বারা দেখা গিয়াছে যে দুইটি Form পর পর এবং 3 বা 4 ঘণ্টার ব্যবধানে প্রয়োগ করিলে তাহাদের আত্ম-অনুবন্ধ সমান থাকে। তবে সাধারণতঃ ব্যবহারিক ক্ষেত্রে অভীক্ষা দুইটির মধ্যে 15 দিন হইতে 30 দিনের ব্যবধান রাখা উচিত।

(গ) Split-Half পদ্ধতি :

এই পদ্ধতিতে পূর্ণ অভীক্ষাকে সমান দুইটি সদৃশ অর্ধে বিভক্ত করিয়া তাহাদের ফলের অনুবন্ধ নির্ণয় করা হয়। এই অর্ধ অভীক্ষার অনুবন্ধ হইতে সমগ্র অভীক্ষাটির আত্ম-অনুবন্ধ, Spearman Brown-এর সূত্র (47 নং) হইতে বাহির করা যায়। প্রথম হইতে একটি অন্তর একটি প্রশ্নের ফল লইয়া দুই সারি নম্বরের অনুবন্ধ নির্ণয় করিতে হয় এবং পরে গোটা অভীক্ষার আত্ম-অনুবন্ধ নির্ণয় করিতে হয়। যেমন—অভীক্ষার 1, 3, 5, 7, 9 ইত্যাদি এবং 2, 4, 6, 8, 10 ইত্যাদি প্রশ্নে বিভিন্ন ছাত্রের ফল পাশাপাশি রাখিতে হইবে। অভীক্ষাকে দুইটি অর্ধে বিভক্ত করিবার বিভিন্ন পদ্ধতি থাকিলেও উপরের যুগ্ম-বিযুগ্ম পদ্ধতি অধিকাংশ ক্ষেত্রে ব্যবহৃত হয়।

$$r_{tt} = \frac{2r_{hh}}{1+r_{hh}} \qquad \text{..... ....(47)}$$

যখন r_{hh} অর্ধ অভীক্ষার আত্ম-অনুবন্ধ-সহ এবং r_{tt} অভীক্ষার আত্ম-অনুবন্ধ সহগ।

উপরের সূত্রের সাহায্যে সমগ্র অভীক্ষার আত্ম-অনুবন্ধ-সহগ বাহির করিবার সময় লক্ষ্য রাখিতে হইবে যে ঐ দুই অর্ধের মাঝে যেন ভর-সামঞ্জস্য বর্তমান থাকে, অর্থাৎ তাহাদের Mean, SD, Skewness প্রভৃতি সমান এবং বিশেষ করিয়া এই দুই অর্ধের বিষয় বস্তু যেন সদৃশ থাকে। নতুবা সমগ্র অভীক্ষার আত্ম-অনুবন্ধ-সহগ সঠিক তথ্য পরিবেশন করিবে না।

যখন অর্ধ-অভীক্ষার আত্ম-অনুবন্ধ-সহগ অর্থাৎ $r_{hh} = \cdot 80$, তখন 47 নং সূত্রানুযায়ী সমগ্র অভীক্ষার আত্ম-অনুবন্ধ সহগ বা r_{tt} হইবে $\frac{2 \times \cdot 80}{1 + \cdot 80}$ বা ·89. সুতরাং দেখা যাইতেছে যে r_{tt} সব সময়েই r_{hh} হইতে বড় হইবে।

অনেক সময় সমান্তরাল প্রশ্ন গুচ্ছ প্রস্তুত বা একই অভীক্ষার পুনর্বার প্রয়োগ সম্ভবপর নয়। Performance test যেমন ছবি পূরণ, ধাঁধাঁ সমাধান, form board সাজান বা ব্যক্তিত্ব, কোন বিষয়ে প্রবণতা, আগ্রহ নিরূপণ প্রভৃতি বহু

অভীক্ষাতে সমান্তরাল অভীক্ষা বা একই অভীক্ষা পুনর্বার প্রয়োগ সহজ সাধ্য নয়। সেই সকল সময়ে Split-Half পদ্ধতি একান্ত অপরিহার্য। কারণ সময়ক্ষেপের দরুন পরীক্ষার্থীর আগ্রহ, উৎসাহ বা মানসিক কাঠামো অনেকাংশই পরিবর্তিত হয়।

এখন প্রশ্ন উঠিতে পারে যে বহু প্রকারে অভীক্ষাকে দুই অর্ধে ভাগ করা যাইতে পারে। সুতরাং সহগ এক না হইয়া বহু হইবে। যখন item-গুলির একই কাঠিন্য থাকিবে, কোন বেগের (speed) অভীক্ষা প্রয়োগ করা হইবে বা ব্যক্তিত্ব পরীক্ষা করা হইবে—তখন উক্ত সহগ বিভিন্ন হইবে। এখন কোন সহগটিকে ঠিক বলিয়া ধরা হইবে এবং কেন,—তাহা বিচার করা কঠিন। তবে মানসিক শক্তির (power) পরীক্ষায় item-গুলি কাঠিন্যের ক্রমানুসারে সাজান থাকে। তাই সেখানে যুগ্ম-বিযুগ্ম পদ্ধতি অভীক্ষার নির্ভরযোগ্যতা নিরূপণে সমর্থ।

(ঘ) Rational Equivalence পদ্ধতি ঃ

উল্লিখিত পদ্ধতি তিনটিতে যে সকল অসুবিধার কথা বলা হইয়াছে তাহা দূর করিবার জন্য Rational Equivalence পদ্ধতি গ্রহণ করা যাইতে পারে। এখানে দুইটি প্রশ্নগুচ্ছ A-a, B-b, C-c ইত্যাদি এইভাবে লওয়া হয় যে তাহারা পরস্পর স্থান বিনিময় করিতে পারে এবং প্রতি item-এর সহিত তাহার সদৃশ item-এর অনুবন্ধ সহগও সদৃশ। তাই এই সদৃশ অভীক্ষা এবং স্থান পরিবর্তনোপযোগী বৈশিষ্ট্য এই পদ্ধতিতে আছে বলিয়া ইহা equivalent. এই পদ্ধতি অভীক্ষার item-এ item-এ অনুবন্ধ এবং একটি item-এর সহিত সমগ্র অভীক্ষার অনুবন্ধের উপর জোর দেয়। নীচের 48নং সূত্রানুযায়ী reliability সহগ বিশ্বস্তভাবে বাহির করা যায়।

$$r_{11} = \frac{n}{n-1} \times \frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2} \qquad \ldots\ldots (48)$$

যেখানে r_{11} = সমগ্র অভীক্ষার reliability সহগ

n = অভীক্ষার মোট প্রশ্ন সংখ্যা

σ_t = সমগ্র নম্বরের SD

p = কোন প্রশ্ন নির্ভুলভাবে উত্তর করিয়াছে এমন ছাত্রের অনুপাত

এবং $q = (1-p)$ = ভুল উত্তর করিয়াছে এমন ছাত্রের অনুপাত।

(C) Reliability সম্পর্কে সতর্কতা ঃ

অনুবন্ধ সহগের ন্যায় Reliability সহগ সম্পর্কে কয়েকটি সতর্কতা ও বৈশিষ্ট্যের আলোচনা প্রয়োজন। যে সকল শর্তের উপর নির্ভর করিয়া উহা পাওয়া যায়—সেই সকল শর্তের উপর ইহার পরিমাণ নির্ভরশীল। নীচে কয়েকটি শর্ত আলোচিত হইল।

(ক) অভীক্ষা প্রযুক্ত বিভিন্ন গোষ্ঠীর উপর reliability নির্ভর করে। গোষ্ঠী যতই কোন ঠাসা হইবে r_{tt}-এর মান ততই ক্ষুদ্র হইবে। আমরা জানি যে $r_{tt}=1-\sigma_e^2/\sigma_t^2$. এখন Error variance যদি অপরিবর্তিত থাকে σ_t-এর মান ক্ষুদ্র হইলে r_{tt}-এর মানও ক্ষুদ্র হইবে।

(খ) দ্বিতীয় শর্ত হইল অভীক্ষার দৈর্ঘ্য। Split-Half পদ্ধতিতে পূর্বেই বলা হইয়াছে যে অর্ধ অভীক্ষা অপেক্ষা সমগ্র অভীক্ষা বেশী reliability সহগ প্রদান করে। তাহাতেই বুঝা যাইতেছে যে প্রশ্ন সংখ্যা বেশী হইলে সহগও বেশী পাওয়া যাইবে। সেইজন্য একই ধরন বা প্রকৃতি বিশিষ্ট অতিরিক্ত প্রশ্ন অভীক্ষাতে ভর্তি করা যাইতে পারে। একই ধরনের বলিতে এই বুঝান হইতেছে যে অভীক্ষার নিজস্ব আত্ম-অনুবন্ধের ন্যায় এই অতিরিক্ত প্রশ্নগুলির সহিত অভীক্ষার সকল প্রশ্নের সেইরূপ অনুবন্ধ রক্ষিত হইবে এবং একই কাঠিন্যের ক্রমানুসারে সজ্জিত থাকিবে। অভীক্ষার প্রশ্ন সংখ্যা n-গুণ বর্ধিত হইলে Spearman-Brown-এর সূত্রানুযায়ী r_{tt} বাহির করা হইয়া থাকে। আবার বিপরীতক্রমে একটি নির্দিষ্ট r_{tt} পাওয়ার জন্য অভীক্ষার প্রশ্ন সংখ্যাকে কতগুণ বাড়াইতে হইবে—তাহাও উক্ত সূত্র দ্বারা নির্ণয় করা হয়।

(গ) অভীক্ষার ফলের ন্যায় Rating এবং অন্যান্য বিচারের reliability নির্ণয় পূর্ববর্ণিত পরিসংখ্যান দ্বারা সম্ভবপর। এখানে বিচারকের পরিমাপকে পদমর্যাদানুযায়ী বা rate অনুসারে সাজান হয়। একই বিচারকের কিছু সময় ব্যবধানে দুইবার পরিমাপ বা বর্তনী পদ্ধতিতে বিভিন্ন বিচারকের পরিমাপের মধ্যে অনুবন্ধ নির্ণয় করিয়া সহগ নির্ণয় করা যায়। তবে মনে রাখা দরকার যে, যেভাবে ঐ সহগগুলি পাওয়া যায় তাহারই পটভূমিকায় সে গুলির অর্থ নির্ণয় বা সমালোচনা করা উচিত।

B. পরিমাপের যাথার্থতা ঃ

(ক) কোন অভীক্ষা যাহা পরিমাপ করিতে ইচ্ছুক কেবল তাহা পরিমাপ করিলে, সেই অভীক্ষা যাথার্থ (valid) বলিয়া পরিগণিত হয়। স্মৃতির অভীক্ষা

কেবল নিছক স্মৃতি, বুদ্ধির অভীক্ষা নিছক বুদ্ধির পরিমাপ করিলে অভীক্ষাকে যাথার্থ বলা হয়। ইংরেজী ভাষার আবরণে যদি অংকের অভীক্ষা আচ্ছাদিত থাকে—যাহার ফলে অর্থ বুঝিবার অক্ষমতায় ছাত্ররা জানা সত্ত্বেও অংকের উত্তর করিতে অসমর্থ হয়—তাহা হইলে সেই অংকের অভীক্ষাটি যাথার্থ নয়।

আবার ব্যক্তিত্ব পরিমাপের সময় বিশ্লেষিত গুণগুলির জন্য ভিন্ন ভিন্ন প্রশ্নাবলীর দরকার হয়। কিন্তু প্রতি গুণের মধ্যে লুকায়িত বিভিন্ন ভেজাল গুণগুলির পরিমাপকে প্রভাবান্বিত করে।

যাথার্থ অভীক্ষা ভবিষ্যৎবাণীর সহায়তা করে। ভর্তির সময়ে প্রযুক্ত অভীক্ষার ফলের সহিত বৎসরের শেষের পরীক্ষার ফলের অনুবন্ধ-সহগ যদি উচ্চ হয় তবে বুঝিতে হইবে যে প্রথমোক্ত অভীক্ষাটি যাথার্থ এবং উহা ভবিষ্যৎবাণী করিতে সক্ষম। অনেক সময় বুদ্ধি বা অন্যান্য অভীক্ষা—কেরাণীর কাজ, সৈনিকের কাজ বা কোন শিল্পের কাজের জন্য ভবিষ্যৎবাণী করিতে পারে। সুতরাং কোন অভীক্ষার যাথার্থতাই তাহার ভবিষ্যৎবাণী করার ক্ষমতা।

(খ) যাথার্থতা বাহির করার সর্বাপেক্ষা কঠিন পথ হইতেছে যে আমরা যাহা পরিমাপ করিতে যাইতেছি তাহার উপযুক্ত criteria বাহির করা। কারণ মানসিক অভীক্ষাতে এই ধরনের সঠিক criteria-র সন্ধান পাওয়া খুবই কষ্টকর।

যাথার্থতা অভীক্ষার একটি আপেক্ষিক গুণ। কোন একটি বিশেষ উদ্দেশ্যে বা অবস্থায় কোন অভীক্ষা যাথার্থ হইতে পারে—সব অবস্থাতে নয়।

(গ) যাথার্থতা নির্ণয় করার পদ্ধতি

বিদ্যালয়ে বিভিন্ন শিক্ষাগত বিষয়ের Standard অভীক্ষা—বিভিন্ন বিষয়ে কোন্ বয়সের ছাত্রের কতটা জ্ঞান থাকা দরকার—শিক্ষাবিদ্‌কে তাহা বুঝাইতে সাহায্য করে। যখন Sampling দ্বারা উপযুক্ত অভীক্ষা বাছিয়া প্রতিভূ ছাত্র/ছাত্রীর উপর প্রয়োগ করা হয় এবং উপযুক্ত দক্ষ বিচারক দ্বারা তাহাদের ফলাফলের বিচার করা হয়—তখনই কোন অভীক্ষার যাথার্থতা পাওয়ার সম্ভাবনা থাকে।

পরীক্ষার মাধ্যমে কোন অভীক্ষার যাথার্থতা নির্ণয় করা যায়। কোন স্বাধীন criterion-এর সংগে অভীক্ষার অনুবন্ধ যদি উচ্চ হয়, তবে অভীক্ষাটি যাথার্থ বলিয়া গ্রহণ করা হয়। কিন্তু এই স্বাধীন criterion বাহির করা কষ্টকর। অনেক

সময় বিদ্যালয়ে শ্রেণীর জ্ঞানের পরিমাপ, শিক্ষক-নির্ধারিত দক্ষতার পরিমাপ বা অন্যান্য ক্ষমতার সহিত তুলনা করিয়া বুদ্ধির অভীক্ষার যাথার্থতা নিরূপণ করা হয়। তেমনই সময়ের সংগে তুলনা করিয়া কোন বাণিজ্যিক অভীক্ষার যাথার্থতা যাচাই হইতে পারে। তবে মনে রাখিতে হইবে যে কোন অভীক্ষা এবং criterion-এর মধ্যে উচ্চ অনুবন্ধ থাকিলে ঐ অভীক্ষাটি যাথার্থ বলিয়া বিবেচিত হইবে—যদি (১) criterion-টি স্বাধীন হয়, এবং (২) criterion-টি নির্ভরযোগ্য হয়।

TABLES

* * * *

A. Areas, Normal Probability Curve
B. Ordinates of the Normal Probability Curve
E. χ^2—Table
G. Calculation of T Scores
J. Cosines of Angles (for computing r_t)

TABLE A Fractional parts of the total area (taken as 10,000) under the normal probability curve, corresponding to distances on the base-line between the mean and successive points laid off from the mean in units of standard deviation

Example: between the mean and a point $1.38\sigma \left(\frac{x}{\sigma} = 1.38\right)$ are found 41.62% of the entire area under the curve.

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3290	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3.0	4986.5	4986.9	4987.4	4987.8	4988.2	4988.6	4988.9	4989.3	4989.7	4990.0
3.1	4990.3	4990.6	4991.0	4991.3	4991.6	4991.8	4992.1	4992.4	4992.6	4992.9
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

TABLE B Ordinates of the normal probability curve expressed as fractional parts of the mean ordinate, y_o

The height of the ordinate erected at the mean can be computed from $y_o = \frac{N}{\sigma\sqrt{2\pi}}$ where $\sqrt{2\pi} = 2.51$ and $\frac{1}{\sqrt{2\pi}} = .3989$.

$\frac{x}{\sigma}$	0	1	2	3	4	5	6	7	8	9
0.0	100000	99995	99980	99955	99920	99875	99820	99755	99685	99596
0.1	99501	99396	99283	99158	99025	98881	98728	98565	98393	98211
0.2	98020	97819	97609	97390	97161	96923	96676	96420	96156	95882
0.3	95600	95309	95010	94702	94387	94055	93723	93382	93024	92677
0.4	92312	91399	91558	91169	90774	90371	89961	89543	89119	88688
0.5	88250	87805	87353	86896	86432	85962	85488	85006	84519	84060
0.6	83527	83023	82514	82010	81481	80957	80429	79896	79359	78817
0.7	78270	77721	77167	76610	76048	75484	74916	74342	73769	73193
0.8	72615	72033	71448	70861	70272	69681	69087	68493	67896	67298
0.9	66689	66097	65494	64891	64287	63683	63077	62472	61865	61259
1.0	60653	60047	59440	58834	58228	57623	57017	56414	55810	55209
1.1	54607	54007	53409	52812	52214	51620	51027	50437	49848	49260
1.2	48675	48092	47511	46933	46357	45783	45212	44644	44078	43516
1.3	42956	42399	41845	41294	40747	40202	39661	39123	38569	38058
1.4	37531	37007	36487	35971	35459	34950	34445	33944	33447	32954
1.5	32465	31980	31500	31023	30550	30082	29618	29158	28702	28251
1.6	27804	27361	26923	26489	26059	25634	25213	24797	24385	23978
1.7	23575	23176	22782	22392	22008	21627	21251	20879	20511	20148
1.8	19790	19436	19086	18741	18400	18064	17732	17404	17081	16762
1.9	16448	16137	15831	15530	15232	14939	14650	14364	14083	13806
2.0	13534	13265	13000	12740	12483	12230	11981	11737	11496	11259
2.1	11025	10795	10570	10347	10129	09914	09702	09495	09290	09090
2.2	08892	08698	08507	08320	08136	07956	07778	07604	07433	07265
2.3	07100	06939	06780	06624	06471	06321	06174	06029	05888	05750
2.4	05614	05481	05350	05222	05096	04973	04852	04734	04618	04505
2.5	04394	04285	04179	04074	03972	03873	03775	03680	03586	03494
2.6	03405	03317	03232	03148	03066	02986	02908	02831	02757	02684
2.7	02612	02542	02474	02408	02343	02280	02218	02157	02098	02040
2.8	01984	01929	01876	01823	01772	01723	01674	01627	01581	01536
2.9	01492	01449	01408	01367	01328	01288	01252	01215	01179	01145
3.0	01111	00819	00598	00432	00309	00219	00153	00106	00073	00050
4.0	00034	00022	00015	00010	00006	00004	00003	00002	00001	00001
5.0	00000									

TABLE E χ^2 Table.

df	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0.00393	0.0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345
4	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	3.325	4.168	5.380	6.303	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

TABLE G To facilitate the calculation of *T* scores

The percents refer to the percentage of the total frequency below a given score +1/2 of the frequency on that score. *T* scores are read directly from the given percentages.*

Percent	*T* score	Percent	*T* score
.0032	10	53.98	51
.0048	11	57.93	52
.007	12	61.79	53
.011	13	65.54	54
.016	14	69.15	55
.023	15	72.57	56
.034	16	75.80	57
.048	17	78.81	58
.069	18	81.59	59
.097	19	84.13	60
.13	20	86.43	61
.19	21	88.49	62
.26	22	90.32	63
.35	23	91.92	64
.47	24	93.32	65
.62	25	94.52	66
.82	26	95.54	67
1.07	27	96.41	68
1.39	28	97.13	69
1.79	29	97.72	70
2.28	30	98.21	71
2.87	31	98.61	72
3.59	32	98.93	73
4.46	33	99.18	74
5.48	34	99.38	75
6.68	35	99.53	76
8.08	36	99.65	77
9.68	37	99.74	78
11.51	38	99.81	79
13.57	39	99.865	80
15.87	40	99.903	81
18.41	41	99.931	82
21.19	42	99.952	83
24.20	43	99.966	84
27.43	44	99.977	85
30.85	45	99.984	86
34.46	46	99.9890	87
38.21	47	99.9928	88
42.07	48	99.9952	89
46.02	49	99.9968	90
50.00	50		

* *T* scores under 10 or above 90 differ so slightly that they cannot be read as different two-place numbers.

TABLE J Values of r_t taken as the cosine of an angle

Angle	Cosine	Angle	Cosine	Angle	Cosine
0°	1.000	41°	.755	73°	.292
5	.996	42	.743	74	.276
10	.985	43	.731	75	.259
15	.966	44	.719	76	.242
16	.961	45	.707	77	.225
17	.956	46	.695	78	.208
18	.951	47	.682	79	.191
19	.946	48	.669	80	.174
20	.940	49	.656	81	.156
21	.934	50	.643	82	.139
22	.927	51	629	83	.122
23	.921	52	.616	84	.105
24	.914	53	.602	85	.087
25	.906	54	.588	90	.000
26	.899	55	.574		
27	.891	56	.559		
28	.883	57	.545		
29	.875	58	.530		
30	.866	59	.515		
31	.857	60	.500		
32	.848	61	.485		
33	.839	62	.469		
34	.829	63	.454		
35	.819	64	.438		
36	.809	65	.423		
37	.799	66	.407		
38	.788	67	.391		
39	.777	68	.375		
40	.766	69	.358		
		70	.342		
		71	.326		
		72	.309		

প্রাসঙ্গিক গ্রন্থ

* * * *

[নিম্নলিখিত পুস্তকগুলি গ্রন্থকারকে সাহায্য করিয়াছে। শিক্ষা এবং মনস্তত্ত্বের শিক্ষার্থীদের বিশেষ পাঠের জন্য এই বইগুলি বিশেষ উপকারে লাগিবে।]

Edwards, A. L. Statistical Analysis. New York : Rinehart & Co., 1946.

Fisher, R. A. Statistical Methods for Research workers. New York, 1950.

Garrett, H. E. Statistics in Psychology and Education. Longmans, Green and Co., 1959.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York : Me Graw-Hill Book Co., 1956.

Jones, D. C., A First course in Statistics. London. G. Bell and Sons, 1921.

Linquist, E. F. Statistical Analysis in Educational Research. Boston : Houghton Mifflin Co., 1940.

Mc Nemar, Q. Psychological Statistics. New York : John wiley and Sons, 1955.

Tate, M. W. Statistics in Education. New York : Macmillan Co., 1955.

Walker, Helen M. Elementary Statistical Methods. New York : Henry Holt and Co., 1943

Yule, G. U., An introduction to the Theory of statistics. London, C. Griffin & Co., 1932.